# NAVAL POSTGRADUATE SCHOOL

## MONTEREY, CALIFORNIA

# THESIS

**TESTING AND DEMONSTRATING SPEAKER VERIFICATION TECHNOLOGY IN IRAQI-ARABIC AS PART OF THE IRAQI ENROLLMENT VIA VOICE AUTHENTICATION PROJECT (IEVAP) IN SUPPORT OF THE GLOBAL WAR ON TERRORISM (GWOT)**

by

Jeffrey W. Withee
Edwin D. Pena

September 2007

| | |
|---|---|
| Thesis Advisor: | James F. Ehlert |
| Thesis Co-Advisor: | Pat Sankar |

**Approved for public release; distribution is unlimited**

THIS PAGE INTENTIONALLY LEFT BLANK

| REPORT DOCUMENTATION PAGE | | *Form Approved OMB No. 0704-0188* |
|---|---|---|

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

| 1. AGENCY USE ONLY *(Leave blank)* | 2. REPORT DATE<br>September 2007 | 3. REPORT TYPE AND DATES COVERED<br>Master's Thesis |
|---|---|---|
| 4. TITLE AND SUBTITLE Testing and Demonstrating Speaker Verification Technology in Iraqi-Arabic as Part of the Iraqi Enrollment Via Voice Authentication Project (IEVAP) in Support of the Global War on Terrorism (GWOT) Security Requirements. | | 5. FUNDING NUMBERS |
| 6. AUTHOR(S) Jeffrey W. Withee; Edwin D. Pena | | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>Naval Postgraduate School<br>Monterey, CA 93943-5000 | | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br>Office of the Secretary of Defense<br>Pentagon, Washington DC 20301-6000 | | 10. SPONSORING/MONITORING AGENCY REPORT NUMBER |

**11. SUPPLEMENTARY NOTES** The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

| 12a. DISTRIBUTION / AVAILABILITY STATEMENT<br>Approved for public release; distribution is unlimited | 12b. DISTRIBUTION CODE |
|---|---|

**13. ABSTRACT (maximum 200 words)**

This thesis documents the findings of an Iraqi-Arabic language test and concept of operations for speaker verification technology as part of the Iraqi Banking System in support of the Iraqi Enrollment via Voice Authentication Project (IEVAP). IEVAP is an Office of the Secretary of Defense (OSD) sponsored research project commissioned to study the feasibility of speaker verification technology in support security requirements of the Global War on Terrorism (GWOT). The intent of this project is to contribute toward the future employment of speech technologies in a variety of coalition military operations by testing speaker verification and automated speech recognition technology in order to improve conditions in the war torn country of Iraq. In this phase of the IEVAP, NPS tested Nuance Inc's Iraqi-Arabic voice authentication application and developed a supporting concept of operations for this technology in support of a new era in Iraqi Banking.

| 14. SUBJECT TERMS Iraq; Speaker verification, voice authentication, voice verification, voice biometrics | 15. NUMBER OF PAGES<br>130 |
|---|---|
| | 16. PRICE CODE |

| 17. SECURITY CLASSIFICATION OF REPORT<br>Unclassified | 18. SECURITY CLASSIFICATION OF THIS PAGE<br>Unclassified | 19. SECURITY CLASSIFICATION OF ABSTRACT<br>Unclassified | 20. LIMITATION OF ABSTRACT<br>UU |
|---|---|---|---|

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

i

THIS PAGE INTENTIONALLY LEFT BLANK

**TESTING AND DEMONSTRATING SPEAKER VERIFICATION TECHNOLOGY IN IRAQI-ARABIC AS PART OF THE IRAQI ENROLLMENT VIA VOICE AUTHENTICATION PROJECT (IEVAP) IN SUPPORT OF THE GLOBAL WAR ON TERRORISM (GWOT)**

Jeffrey W. Withee
Major, United States Marine Corps
B.A., The Citadel, 1996

Edwin D. Pena
Captain, United States Marine Corps
B.A., University of Colorado, 2001

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN INFORMATION TECHNOLOGY MANAGEMENT**

from the

**NAVAL POSTGRADUATE SCHOOL
September 2007**

Authors:         Major Jeffrey W. Withee

                 Captain Edwin D. Pena

Approved by:     James F. Ehlert
                 Thesis Advisor

                 Pat Sankar
                 Thesis Co-Advisor

                 Dan Boger
                 Chairman, Department of Information Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

# ABSTRACT

This thesis documents the findings of an Iraqi-Arabic language test and concept of operations for speaker this verification technology as part of the Iraqi Banking System in support of the Iraqi Enrollment via Voice Authentication Project (IEVAP). IEVAP is an Office of the Secretary of Defense (OSD) sponsored research project commissioned to study the feasibility of speaker verification technology in support security requirements of the Global War on Terrorism (GWOT). The intent of this project is to contribute toward the future employment of speech technologies in a variety of coalition military operations by testing speaker verification and automated speech recognition technology in order to improve conditions in the war torn country of Iraq. In this phase of the IEVAP, NPS tested Nuance Inc.'s Iraqi-Arabic voice authentication application and developed a supporting concept of operations for this technology in support of a new era in Iraqi Banking.

THIS PAGE INTENTIONALLY LEFT BLANK

# TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

# ACKNOWLEDGMENTS

Jeff: Above all else we would like to thank God for this time of fellowship while here at NPS. Additionally, we would like to thank our sponsors at both the Office of the Secretary of Defense and at SPAWAR System Center San Diego, CA for not only providing money, but for providing mentorship as well. We would like to thank our Thesis Advisors for remaining unimpressed and keeping us focused on completing this project. In addition, we would also like to thank Captain Lee, USMC and Major Sipko, USMC, for their excellent work on the IEVAP project prior to Phase 1C. Lastly, I would like to thank my wife, Kara and sons Owen, Angus, and Emmett, for supporting me during this process. "This means" that I could not have done it without you.

Eddie: In addition to those thanked above, we would like to thank Dr. Alex Bordetsky for allowing us to use the CENETIX lab during this experiment and for his mentoring and instruction during our time here at NPS. Additionally, we would like to thank LCDR Jamie Gateau, Eugene Bourakov, and Mike Clement from the CENETIX lab for all of their help and patience with our inane questions. Finally, I would like to thank my wife, Federica, and daughter, Jazmin, for enduring both Jeff and I during this process. Ti amo amore!

THIS PAGE INTENTIONALLY LEFT BLANK

# I. INTRODUCTION

## A. OVERVIEW

This thesis documents the findings of the third part of phase one of the Iraqi Enrollment via Voice Authentication Project (IEVAP Phase 1C). The IEVAP is an Office of the Secretary of Defense (OSD) sponsored research project that studies the feasibility of speaker verification and speech recognition technology in support of security for banking and other security applications primarily in Iraq and for the Global War on Terrorism (GWOT) in general.

Since the toppling of the Baathist regime in 2003, the banking system in Iraq has not improved much from the tribal, cash-based system that existed before the war. This shortcoming has contributed to the inability of the Iraqi government to account for over 12 Billion U.S. dollars during the last four years [1]. As Lieutenant General David H. Petraeus, Commander U.S. Forces Iraq stated in an interview shortly after taking command, "there is no strictly military solution" to this problem in Iraq [2]. If there is to be any hope for stability in Iraq, the problems of corruption, the lack of a banking system, and a lack of information infrastructure (or infostructure) [3] must be addressed at least in parallel but preferably prior to implementing secure financial transaction applications. The system studied for this thesis addresses all of these issues on some level with the following potential benefits:

- Once financial transactions migrate from a cash-based system to an electronic-based system, it will be possible to keep a more accurate record of payments. This will act as both a means of financial accountability as well as a deterrent to corruption by providing evidence for the prosecution of those who attempt embezzlement.

- This technology will provide a secure means to pay Iraqi soldiers and police (such as a debit card system) without having to pay them in cash, which currently leads to a large percentage of the force disappearing for several days while they deliver this cash to their families.

1

- This system can be part of a money-wire transfer system that will decrease the need for travel and the inherent risk that soldiers/police will desert or become victims of robbery, kidnappings, or worse while en route to their villages with cash.

- With decreased corruption, infrastructure improvements will occur at a much lower cost and with a better return on investment for the country.

- This technology can be implemented in security applications at checkpoints for the quick processing of Iraqi VIPs and local nationals.

- In addition, Phase 1A of this research project successfully demonstrated how a voice authentication program could be used to create an appointment system. Such a system would decrease the long lines at military installations, which are prime targets for attack by insurgents.

The vision for this project, once the Proof of Concept (POC) is established and when used in conjunction with other biometric systems and security procedures, speaker verification applications and Automated Speech Recognition (ASR) technologies could become tools for positively identifying individuals in support of the GWOT in a number of different ways. Moreover, IEVAP is an initiative that transcends the potential implementation in Iraq. A successful POC could lead to applications in other stabilization and reconstruction efforts elsewhere, such as in Afghanistan.

In short, this technology should have been considered for operational use at the onset of the redevelopment effort in Iraq, as it may prove imperative for the country's financial stability. The benefits to Iraq are evident and such a system supports the U.S. plan to hand over control of the country to Iraqi nationals and extract its troops from Iraq.

## B. BACKGROUND

OSD tasked the Naval Postgraduate School (NPS) with developing and demonstrating a pilot POC system in support of the IEVAP. The IEVAP is organized into several project phases that are intended to take the POC system from concept development to operational testing in Iraq. This thesis documents the findings of the third sub-phase (Phase 1C) within Phase 1 of the project, which are as follows:

- **Phase 1**. Pilot menu-driven laptop system and demonstration that voice authentication technology can work with sufficient accuracy.

  - **Phase 1A.** Develop and demonstrate a bilingual voice-activated menu-driven phone system in English and Arabic.
  - **Phase 1B.** Test and demonstrate speaker verification technology in English.
  - **Phase 1C.** Test and demonstrate speaker verification technology in Iraqi-Arabic.
- **Phase 2.** Detailed development of enrollment applications

- **Phase 3.** Preparation of systems/applications for deployment

- **Phase 4**. Deployment

- **Phase 5.** Operational testing in Iraq

- **Phase 6.** Broader deployment decision

## C.    RESEARCH QUESTIONS

- Is it possible to create and deploy a phone speaker-verification platform using existing Commercial-Off-The-Shelf (COTS) technologies to assist in security operations and banking application requirements in support of the GWOT?

- What measures must be taken in order to successfully implement this new way of conducting business and mitigating resistance to change?

- In what ways can this technology help stimulate the financial sector in Iraq, while combating corruption and increasing security (concept of operations)?

## D.    SCOPE OF THESIS

This thesis focuses on the technologies addressed in support of Phase 1C of the IEVAP, which includes the development and demonstration of an Iraqi Arabic voice-activated menu-driven telephone system and an analysis of results of the NPS Speaker Verification Test.  The value of this research includes:

- Demonstrating the viability of speaker verification and ASR technology for subsequent research, development, and possible real-world implementation.

- Providing a "quick response" research and development capability to address external customer requirements.

- Selecting the most appropriate hardware, software, and peripherals for a remote demonstration kit (server, voice input devices, etc) for implementing speaker verification and ASR technologies.

## E. RESEARCH METHODOLOGY

This investigation employs the quantitative approach for data collection and analysis. This research consists of the development of an Iraqi Arabic application to assist in combating corruption and securing banking transactions from the Ministerial level on down to the paying of soldiers/police as well as other security applications in Iraq. This research also consists of an analysis of the COTS speaker verification software, Nuance Caller Authentication (NCA) 1.0 for Iraqi-Arabic language.

## F. THESIS ORGANIZATION

Chapter II discusses the technology behind speaker verification. Chapter III is an overview of Nuance Communication, Inc. and its core technologies, operating platform and packaged applications. Chapter IV describes a test to assess the performance of the NCA speaker verification application using the Nuance's Iraqi Arabic language verification master package (language module), to include the identification of equipment (hardware, software and peripherals) used to conduct this test and an analysis of the results of the independent NPS Speaker Verification Test. Chapter V describes the concept of operations and the technical implementation of a telephonic banking system. Chapter VI discusses managing the planned change of the implementation of this system. Finally, Chapter VII concludes with recommendations for possible future work relating to this technology.

# II.    SPEAKER VERIFICATION TECHNOLOGY

## A.    OVERVIEW

The first question that needs to be answered is "why use a biometric authentication for this project?"  Basically, the answer is simple; security is the most important aspect of this project.  The world of security uses three forms of authentication: "something you know—a password, PIN, or piece of personal information (such as your mother's maiden name); something you have—a card key, smart card, or token (like a SecureID card); and/or something you are—a biometric." [4] Out of these three authentication tools, biometrics is the most secure and convenient.  For the most part, biometrics can be neither borrowed, stolen, forgotten, nor forged.  Of course there are always exceptions to the rule, but the victim in one of these rare instances will probably have more to worry about than having someone authenticated in his or her place.  In the specific case of Iraqi Banking, it is very important that transactions occur in an environment of nonrepudiation.  Nonrepudiation is "the ability to ensure that a party to a contract or a communication cannot deny the authenticity of their signature on a document or the sending of a message that they originated" [5]. Simply put if a fraudulent transaction is made, the one who made the transaction cannot deny the fact that he or she made that transaction in question.

## B.    COMPARISON OF VOICE BIOMETRICS

The second question that must be answered is why use "Voice Authentication over other forms of Biometrics?"  The truth is that there are a number of biometrics from which to choose, ranging from Fingerprints, Hand Geometry, Retina, Iris, Face, Signature, and Voice.  Each biometric has both strengths and weaknesses.  Table 1 will help demonstrate why, in this particular case, Voice Authentication is the best tool for the Iraqi Banking System as well as other security problems in Iraq that require controlled access.

| Characteristic | Fingerprints | Hand geometry | Retina | Iris | Face | Signature | Voice |
|---|---|---|---|---|---|---|---|
| Ease of Use | High | High | Low | Medium | Medium | High | High |
| Error incidence | Dryness, dirt, age | Hand injury, age | Glasses | Poor lighting | Lighting, age, glasses, hair | Changing signatures | Noise, colds, weather |
| Accuracy | High | High | Very high | Very high | High | High | High |
| Cost | * | * | * | * | * | * | * |
| User acceptance | Medium | Medium | Medium | Medium | Medium | Very high | High |
| Required security level | High | Medium | High | Very high | Medium | Medium | Medium |
| Long-term stability | High | Medium | High | High | Medium | Medium | Medium |

* The large number of factors involved makes a simple cost comparison impractical.

Table 1.    Comparison of Biometrics [From 4]

In order to fully leverage the information presented through this chart some basic definitions must be given [4]:

### 1.    Ease of Use

This term refers to how much training is required for an individual to use the system.   In this case voice is rated as "high," meaning it has a high ease of use.  A system that is easy to use is very beneficial for this project because the system will need to be accessible to a wide variety of people encompassing both the educated and the uneducated.

### 2.    Error Incidence

This term refers to errors that can affect biometric data.  The two most common are time and environment.  Although the environment will always be a factor, with tuning (greater detail about tuning will be provided in Chapter III) Voice Biometrics can actually improve in accuracy over time.  On the other hand, the human voice can change if an individual suffers from a cold, is under stress, or because of many other various factors.

### 3. Accuracy

Accuracy is the overall ability of the system to allow the right people access and to keep the wrong people out of the system. The two most commonly used methods to rate biometrics are false-accept or false rejection rate. A false-accept is the most dangerous error as it can lead to a greater amount of loss than the false rejection rate. It is important to note that the false rejection rate must also be kept to a minimum to avoid customer dissatisfaction. Although not scored as "very high", voice biometrics, as shown in the results of this research, can still have impressive accuracy.

### 4. Cost

The cost of a system is comprised of many factors ranging from the hardware and software being used to the installation and maintenance required for that hardware and software to be instantiated. Though not featured in Table 1, and even if the unit cost of this entire system is more expensive than the unit cost of other biometric systems, it would still be worth the investment as no additional infrastructure upgrade is required because the system is accessed remotely. Other biometrics do not work remotely, thus requiring a greater number of units to reach more people. It is unlikely that a Voice Biometric System will be more expensive than other biometric systems (since the existing phone lines and wireless communication infrastructures can be used with little or no modifications) and in the long run this type of system has the potential to save money.

### 5. User Acceptance

User acceptance directly relates to how intrusive a biometric is. Although privacy is not a great concern in the middle-east, personal space is of great importance. When searching subjects in Iraq it can quickly be ascertained that they liked neither to be touched nor moved in any way. Because of this issue, many other forms of biometrics are too intrusive for use in Iraq. Voice biometrics, on the other hand, have a high rate of acceptance because all that is required of the user is that he or she be willing to speak. This type of system, therefore, allows for minimal intrusion of personal space.

**6. Required Security**

Required security refers to the level of security at which a biometric should be used. In the case of voice biometrics, the required security is rated as "medium." However, any biometric system including voice biometrics can be configured as a high security system if the situation demands it. Although this particular application will be used primarily for banking, at this point in IEVAP the concern is more for accountability and nonrepudiation than for security.

**7. Long-term Stability**

The long-term stability relates to a biometrics' maturity and standardization throughout the industry. This rating is "medium" in the case of voice biometrics. Automated Speech Recognition (ASR) began in 1920 with the invention of a small toy named Radio Rex who would stand on all four legs when its name was called [6]. But it was not until the 1950s that Bell Labs developed a system that could recognize single digits verbalized with a pause that had a 2% error rate. The 1960s saw continued expansion of this system, but it was not until the 1990s when computing power was such that greater advances and reliability were established.

**8. Other Factors**

Another item of interest is that the technology is such that Speaker Verification lends itself quite well to the mobile environment. This is a huge plus for the environment in Iraq, as many VIPs, such as sheiks and Imams, detest being treated as common or made to wait. In order to ensure that the process is speedy and safe, a Speaker Identification system could be loaded onto a laptop and used remotely as proven in Phase 1A and B of this research project [7]. Such remote access would allow for two important considerations: special treatment for VIPs and as a standoff capability for security personnel. This is a win-win since VIPs do not like to be touched or manhandled in any way. Conversely, security personnel want to be able to authenticate that a person is who they say they are. Without physically engaging a VIP, the security personnel could

simply have them speak into a microphone connected to a laptop. From the gate, security personnel could verify the VIP and allow them the access they require in a quick and non-invasive manner.

## C.      AUTOMATED SPEECH RECOGNITION

Since the advantages of a Speaker Verification System and how it fits this particular task have been discussed, the basics of ASR must now be explored. The subcategory of Voice Recognition has two main areas - Speaker Verification and Speaker Identification. The two are often used interchangeably, but are not one and the same. "Speaker Verification is the process of confirming that a speaker is the person they claim to be; for example, to gain entry to a secure area" [8]. For the IEVAP, speaker verification would be used for gaining access to an account in order to conduct financial transactions. This is not to be confused with Speaker Identification, "the process of determining which speaker in a group of known speakers most closely matches the unknown speaker" [8]. Speaker Identification is primarily used in law enforcement in order to identify if the person is known or unknown.

As mentioned previously, IEVAP focuses on the former, Speaker Verification. In order to successfully use Speaker Verification, the system must combat two types of error: false acceptance and false rejection. False acceptance is when the wrong person, malicious or not, gains access into an account in which he or she is not authorized. False rejection occurs when the right person is rejected from an account into which he or she is authorized to have access. Later in this chapter the balance of these two errors, in terms of rates and how their relationship to each other affects the system as a whole, will be discussed.

## D.      THE PROCESS OF SPEAKER VERIFICATION

There are two things which must be done is order to conduct Speaker Verification: Enrollment and Verification. Both of these processes are not unlike the techniques used for all biometrics. The enrollment process consists of three phases: the capture, the processing and the actual enrollment [9].

9
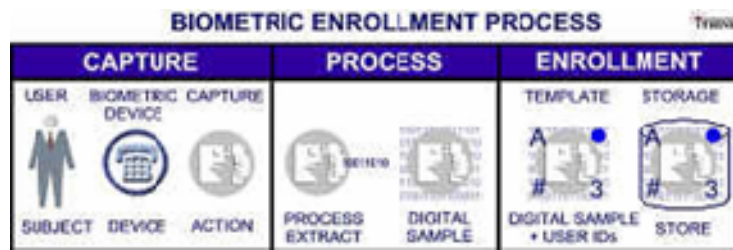
Figure 1.        Biometric Enrollment Process [From 9]

First a user, in this case a speaker, will use a biometric device (such as a cell phone, VOIP, microphone, etc.), and have the voice recorded by a system as a sound file, such as a WAV file. Second, the speaker's voice is processed in order to extract the feature that contains the speaker information and a digital sample is made. From this, the digital sample is paired with an account number or Identification Code which is then stored in a database for use during the verification process. The process of verification is much like the enrollment process.
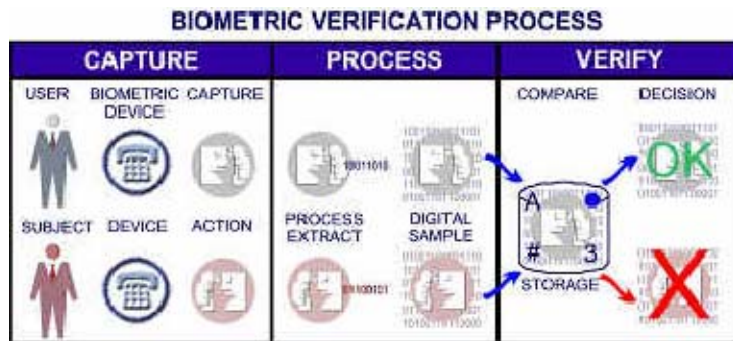


Figure 2.        Biometric Verification Process [From 9]

Again, the speaker's voice is captured using a biometric device and the action is recorded. The speaker's voice is again processed in order to extract the features of the voiceprint and a digital sample is made. Instead of storing that information, the previous information is referenced in order to glean whether or not it is the correct speaker. This is done using a likelihood ratio test to distinguish between the file in the database and the new file that has just been extracted. The system will then generate a ratio or percentage

on the likelihood of the match and compare that ratio to the ratio that meets the threshold of the system. Based on that threshold, the speaker will either be accepted or rejected. The performance measures that are the basis of this acceptance or rejection will be discussed in the next part of this chapter.

## E. PERFORMANCE MEASURES OF BIOMETRICS

When looking at a biometric system, it is important to look at the accuracy rate. That being said, "Asking a system to perform 100% accurately, 100% of the time is clearly unachievable. Machines are prone to inaccuracy, just as the human beings using them are"[10]. The users of a system must look at what is reasonable to the system considering the environment as well as what purpose the biometric is being used for. Therefore, we must examine how the system performs as it pertains to the errors in the system and the overall accuracy of the system.

### 1. Errors

As mentioned previously a Speaker Verification System must deal with two types of Error, False Rejection and False Acceptance. The rate at which these errors occur is a critical part of measuring a systems performance [11]: The false acceptance rate is the probability that an unauthorized individual is authenticated. The false rejection rate is the probability that an authorized individual is inappropriately rejected. The equations provided below calculate both rates:

$$\widehat{FAR} = \frac{\text{number of false acceptances}}{\text{number of impostor attempts}} \qquad (1)$$

$$\widehat{FRR} = \frac{\text{number of false rejections}}{\text{number of enrollee attempts}} \qquad (2)$$

Figure 3.        Equations for False Acceptance and False Rejection Rate [From 11]

The following figure demonstrates the balance between the False Rejection Rates and the False Acceptance Rates using a receiver operating characteristic (ROC) curve. "A ROC Curve is a plot of FAR against FRR for various threshold values for a given application. An example of an ROC Curve is shown in Figure 2, in which the desired area for a given application is at the lower left of the plot, where both types of errors are minimized" [12]. If a system has a high number of false acceptances, it will ultimately have less security. If the system has a high number of false rejections, it will offer less convenience. The following figure demonstrates the difference using a receiver operating characteristic (ROC) curve. The point at which the number of false rejections equals the number of false acceptances is known as the Equal Error Rate (EER).

Figure 4.        Receiver Operating Characteristic Curve [From 12]

Another way to measure accuracy is a variant of the ROC curve known as Detection Error Tradeoff (DET). The DET curve takes the same tradeoff as the ROC curve, but it uses a normal deviate scale. Essentially this takes the same data and moves it away from both the X and Y-axis allowing for greater readability when plotting multiple curves. Figure 5 depicts the two curves side by side [12].
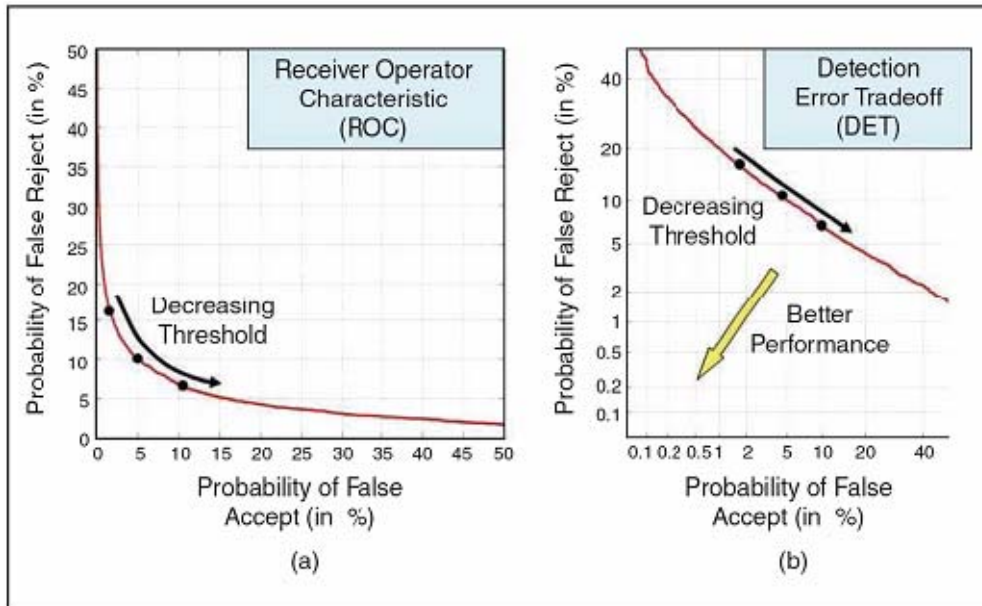
Figure 5.        ROC Curve and DET Curve [From 12]

Remember, these terms refer to the performance of the system, not necessarily with the overall accuracy of the system, although there is a degree of correlation.  The system accuracy has more to do with a single point analysis.

### 2.        Accuracy

As stated previously, accuracy is the ability to keep the wrong people out and let the right people in.  Mathematically, the true accuracy of a system is measured in relation to a single data-point analysis.  In order to get this, the following equation must be used [7]:

$$NT = NTAR + NFRR + NFAR + NTFR.$$

where,

| | |
|---|---|
| NT | The total number of valid verification attempts |
| NTAR | The total number of true accepts |
| NFRR | The total number of false rejects |
| NFAR | The total number of false accepts |
| NTFR | The total number of true failures. |

therefore,

13

Accuracy of the System = ( NT - ( NFRR + NFAR ) ) / NT = ( NTAR + NTFR ) / NT

Note: Nuance presents only FRR and FAR.

### 3.     Confidence Interval

Although a point can give you a good reference for accuracy, it does not reflect the confidence that given the same experiment that these numbers would be the same. Estimating statistical parameters, such as mean or variance from a set of samples, can result in "point estimates." Point estimates are single number estimates of the parameters in question. While very useful in many applications, one limitation of a point estimate is the fact that it conveys no idea of the uncertainty associated with it. If many such point estimates are used in the same analysis, it can become challenging to decipher which estimate is the best/most accurate.

On the other hand, a confidence interval provides a range of numbers (between a lower limit and an upper limit) with a certain degree of probability as to the possible interval of the respective point estimate. Thus, it is easier to conclude that the point estimate with the shortest confidence interval is the most robust and reliable.

### 4.     Statistical Basis

The statistical analysis in the design of the NPS voice verification test was based on the following simplified scenario:

Assume that N speakers, taken at random from the envisaged user population, provide data for the trial. For simplicity, assume also that, for any given trial condition, each speaker makes one verification bid, whose result is either correct or incorrect, and that the results of different speakers' bids are independent. Let the probability of an incorrect verification result for any one bid — that is, the underlying population error rate — be $p$. Then the observed number of errors, $r$, is binomially distributed with mean $Np$ and variance $Np(1-p)$; and the observed error rate $r/N$ has mean $p$ and variance $p(1-p)/N$.

Assuming that the data is "normal," the 05% confidence limit on the observed error rate is expressed as [13]:

$$p \pm 1.96 * \text{sqrt}(( p(1-p)/N)).$$

This equation was computed by measuring 95% of the area, i.e. a 95% probability, on the normal distribution curve, which corresponds to a value of $1.96\sigma$, where $\sigma$ is the standard deviation.

When $p = 0.01$ (or when the population error rate is 1%), the confidence limits are as follows:

$$\pm 1.96*\text{sqrt}((0.0099/N)) = 0.01 \pm 0.195/\text{sqrt}(N)$$

Setting $N$ equal to 1000 gives confidence limits of:

$$0.01 \pm 0.00617 \text{ (i.e. } 1\% \pm 0.617\%) \text{ on the observed error rate.}$$

More accurate estimates of the confidence intervals for small values of $p$ can be derived using the Poisson distribution.

THIS PAGE INTENTIONALLY LEFT BLANK

# III.    NUANCE COMMUNICATIONS, INC.

## A.    BACKGROUND

Nuance Communications, Inc. is a leading, publicly held company (NASDAQ: NUAN) in the development of speech recognition applications.  Company headquarters are in Burlington, Massachusetts but they have expansive complexes throughout the United States.  They also have divisions and training centers in Canada, Latin America (Brazil), Europe (Spain, Italy, France, The Netherlands, Sweden, Hungary, Britain, and Belgium), and Asia (India, South Korea, Australia, Japan, and Hong Kong).   As proof of their unrivaled expertise in the area of speech technology, Nuance was recognized with an unprecedented five awards from Speech Technology Magazine in 2006 for their work in various types of speech technology [14].  Nuance's customers range from banks to government agencies to other businesses that want to integrate speech technology in order to improve customer service while automating personnel intensive applications. Their technology is also being used for increased productivity, convenience in applications such as dictation, transcribing, voice activated calling, and voice activated selection of music for MP3 players.  Some of their clients include: AT&T Wireless, Sprint PCS, T-Mobile, Japan Telecom, Banco Bradesco, British Airways, Charles Schwab, Merrill Lynch, General Motor's OnStar and United Parcel Services [15].  In 2005, Nuance and ScanSoft (another industry leader in voice Interfaces and document management) merged and retained the Nuance name [16].

## B.    CORE TECHNOLOGIES

The following is a general overview of Nuance's core technologies, platform and packaged applications. The information provided below was gathered from datasheets that are readily accessible from Nuance's website at http://www.nuance.com/news/datasheets/ .

Nuance's core technologies in speech consist of three primary applications: speech recognition, text-to-speech, and speaker verification that enable recognition and

understanding of simple responses and complex conversational requests, the conversion of written information into speech, and the authentication of an individual's identity.

This phase of the experiment used Nuance Recognizer 8.5.  In April 2007, Nuance launched version 9.0 that improved the decoder but mostly uses components from ScanSoft's Openspeech Recognizer 3 and Nuance's Recognizer 8.5.  Nuance claims that version 9.0 will give significant improvements over past iterations of their recognizer software.  Below is an illustration of the recognizer process as well as a chart with some of the improvement claims made by Nuance:
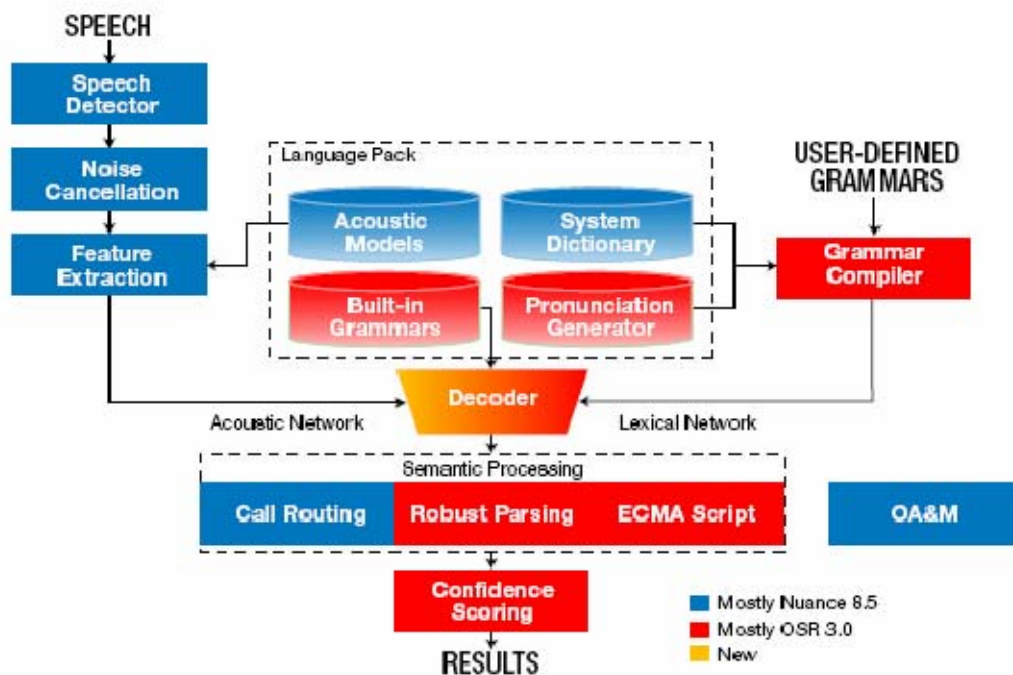


Figure 6.       Nuance Recognizer combines elements of OpenSpeech Recognizer 3 and Nuance 8.5 [From 17]

| Language | Achieved RERR% vs. OSR3 | Achieved RERR% vs. Nuance 8.5 |
|---|---|---|
| U.S. English | 27% | 26% |
| Australian English | 35% | 29% |
| UK English | 15% | 32% |
| German | 33% | 16% |
| Canadian French | 27% | 39% |
| French | 14% | N/A |
| Spanish | 45% | N/A |
| Indian English | 27% | N/A |

Table 2.    Relative Error Rate Reduction (RERR) for Nuance Recognizer, from internal Nuance benchmark testing.  Results represent averages across multiple recognition tasks such as digit strings, alphanumeric spellings, and item lists such as stocks or city names [From 17]

Some of Nuance Recognizer's key features include support for simultaneous load balancing and fault tolerance across speech recognition, speaker verification and text-to-speech operations.  These solutions ensure efficient use of system resources.  Among the 44 languages and dialects that Nuance Recognizer supports are American English, Australian/New Zealand English, Canadian French, Cantonese, European French, German, Italian, Japanese, Jordanian Arabic, Mandarin, Portuguese, Spanish, Swedish and UK English.  For the purposes of this proof of concept, Nuance developed the grammar and models for Iraqi Arabic using native Iraqi speakers now living in Jordan.  Below are some of the additional advanced features available with Nuance Recognizer:

- **Say Anything**TM is a feature that includes Nuance's statistical language models (SLM) and robust natural language interpretation (robust NL) technologies.  It enables automation of complex and open-ended dialogues that are difficult or impossible to implement using traditional grammars.

- **Listen & Learn**TM is a task adaptation feature.  Task adaptation is a self-tuning feature of the Nuance System that automatically improves recognition performance of deployed applications.  Because of this feature, performance will actually improve as more utterances are recorded.

19

- **AccuBurst**™ is a dynamic accuracy feature that allows the recognizer to trade off accuracy against speed according to the load of the machine on which it is running. With dynamic accuracy turned on, the system uses resources when they are available. The recognition rate is then improved during non-busy hours without any noticeable slowdown for the user.

### 1. Text-to-Speech

Nuance Vocalizer 4.0 delivers voice-enabled dynamic and frequently changing information through a phone or other audio system in a natural sounding voice. Because it converts text to speech, there is less of a need to rerecord information that changes often so long as the word components of the desired phrase have already been recorded. This reduces costs in one of the most expensive aspects of speech technology, voice talent. Nuance Vocalizer currently offers 18 languages and a limited amount of speech in Iraqi Arabic for the purposes of this experiment.

### 2. Speaker Verification

Nuance Verifier 3.5 is one of the key features of this technology and what really sets Nuance apart from its competitors. Some of the features Nuance Verifier offers include [18]:

- Effective in a wide range of environments—landline, wireless or hands free phones.

- One-time enrollment for verification during any subsequent call, from any type of phone.

- Speaker identification allows multiple users to share [the same] account or identifier.

- Ongoing adaptation of voiceprint characteristics as voices change or age, improving the quality of voiceprints for faster verification.

- Supports random prompting to safeguard against recording.

- Integration of verification and speech recognition that combines "who you are" with "what you know" in a single phrase.

20

- Unique combination of voice authentication and speech recognition delivers multi-factor security (knowledge verification and voice authentication).

- Verification using letters, numbers, alphanumeric strings, phrases, etc.

- Dynamically detects if more information is needed to verify callers.

- Advanced logging for more effective application tuning.

- Extensive language support.

- Can increase system automation and cost savings by reducing reliance on live agents to identify customers.

- Can reduce occurrences of PIN resets, reducing call center costs.

- Can increase security of information access, reducing the potential for fraud and identity theft.

- Can improve customer service with a convenient means of security.

- Voiceprint storage is nearly impossible to "reverse engineer" for application access.

- Flexible means of verification for individuals or groups.

- Simple maintenance, load balancing and fault tolerance.

## C.  VOICE PLATFORM

Nuance's Voice Platform (NVP) 3.0 ties in the three core technologies previously discussed.  This platform is the foundation on which voice applications are developed and deployed.  It is the link between the user and the backend system that the user wants to access.  NVP 3.0 is based upon open standards and the Voice Extensible Markup Language (VoiceXML) 2.0 standard.  VoiceXML 2.0 is the current international standard

developed by World Wide Web Consortium (W3C) VoiceXML Forum.  Unlike other systems that are based on legacy touch-tone systems and proprietary standards, NVP 3.0 uses open standards that allow developers to use the best and newest features and technologies available in voice applications.  The Voice Platform is comprised of four functional areas: Nuance Conversation Server, Nuance Application Environment, Nuance CTI Gateway, and Nuance Management Station.
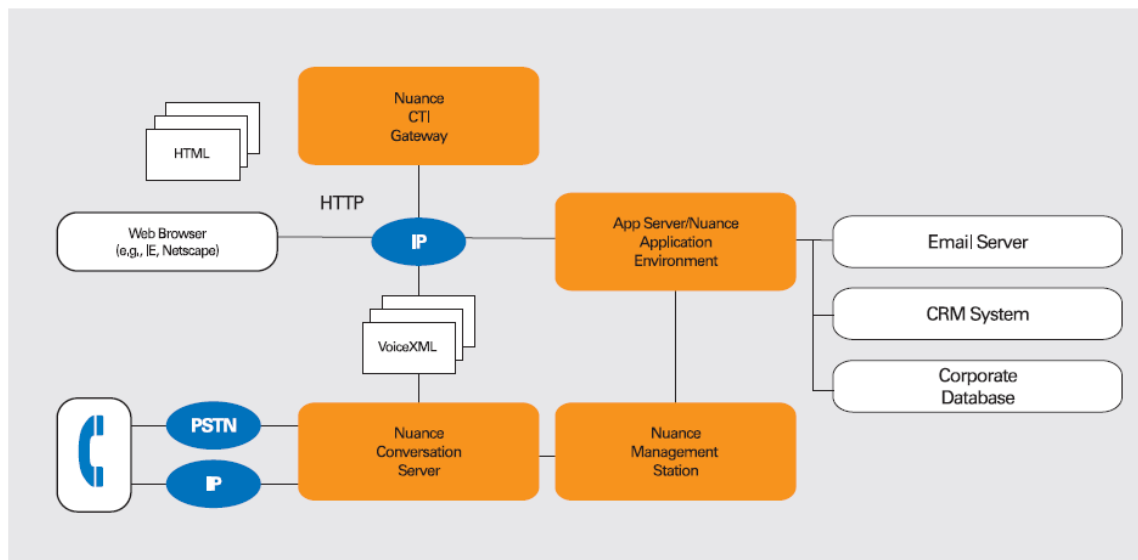


Figure 7.        Overview of NVP 3.0 and its functional areas [From 18]

The following is from a Nuance Datasheet on Voice Platform 3.0:

- The **Nuance Conversation Server** includes a VoiceXML Interpreter integrated with Nuance's speech recognition, text-to-speech and voice authentication technologies.  Using standard Internet protocols, the Nuance Conversation Server fetches VoiceXML applications generated by the Nuance Application Environment or other application frameworks. The Nuance Conversation Server also provides the interfaces to the telephony network via support for commercial-off-the-shelf (COTS) telephony network interface cards or through support for Voice over Internet Protocol (VoIP) through Session Initiated Protocol (SIP).

- The **Management Station** provides an intuitive graphical user interface (GUI) for configuring, deploying, administering, and managing voice applications. It also provides centralized management of the services on the Conversation Server hosts. The three main functions of the management station are System Management and Control, System Performance Analysis and Data Management.

- The **Nuance Application Environment** (NAE) is an integrated graphical application development and runtime environment that facilitates the design, development, deployment, and maintenance of speech applications. This framework can run on widely used application servers to create dynamically generated VoiceXML applications. The voice application can readily integrate to a broad range of backend databases, applications, and legacy systems using web services standards and a variety of pre-packaged interfaces offered by application server vendors. Application developers can also analyze and tune voice application performance and usability. Additionally, a key feature of NAE is that it is an intuitive development environment that enables reusability of application modules.

- The **Nuance Computer Telephony Integration** (**CTI**) **Gateway** provides packaged integrations to leading CTI servers. NVP 3.0 can be integrated into CTI environments from leading vendors such as Aspect, Cisco, and Genesys, allowing enterprises to deploy a best-of-breed, integrated contact center solution that can provide callers with a consistent, high-quality user experience [19].

## D. PACKAGED SPEECH APPLICATIONS

Among the numerous voice enabled applications available from Nuance, a final one that is worth mentioning is Nuance Caller Authentication (NCA) 1.0 [7] NCA 1.0 is a packaged application that can get an organization up and running quickly since it has most of the desired features of speaker recognition and authentication already built in. Using NCA allows for a more advanced level of security than legacy systems that use knowledge questions or DTMF input of PINs. This application is no longer sold as a package by Nuance, but you can order what amounts to the same application through Nuance's custom application order process. Nuance has a very diverse application lineup to address the voice-enabled application needs of any business, state or government agency. More information is available on their website: www.nuance.com.

THIS PAGE INTENTIONALLY LEFT BLANK

# IV.   SPEAKER VERIFICATION TEST

## A.   OVERVIEW

The purpose of the Independent NPS Speaker Verification Test was to validate the accuracy claims of Nuance's speaker verification technology and their test with native Iraqi Arabic speakers residing in Jordan.  Having been granted sole-source justification to hire Nuance, Nuance conducted a 200-person Iraqi Arabic speaker verification test; for details of the Nuance test, please refer to Appendix A.  NPS's Independent Test was conducted using 45 native Iraqi speakers now residing in California.  The comparison of the two tests was made using the performance measures of false reject rate (FRR) and false accept rate (FAR).  The test was conducted using Nuance's packaged speaker verification application, Nuance Caller Authentication (NCA) 1.0, using their Iraqi Arabic Language Verification Package.  Powered by Nuance's Verifier, NCA uses voice biometric technology to capture the physical and behavioral characteristics of the human voice in a voice model.  After associating a particular voice with an account number, it will only allow access to that account if it believes the requesting voice is the original voice within a predetermined confidence percentage.

## B.   EQUIPMENT LIST

For the Independent NPS test, the following hardware, software, and peripherals were used:

### 1.   Hardware

Based on Nuance's software requirements, NPS purchased or borrowed the following hardware in order to conduct this test.

- HP xw9300 workstation
- (2) AMD Opteron™ Processor 246 (1.99 GHz each)
- 2 GB DDR2-533 SDRAM
- (2) 100GB Hard Drives

Figure 8.　　　HP xw9300 workstation (Beaker)

This server, affectionately known as "Beaker," was chosen for its processing power, memory capability, and because it already existed on the school network. Nuance recommended (at a minimum) using a 2 GHz processor with 2 GB RAM on a Microsoft Windows XP based system.  In distributed architectures, the minimum requirement is 3 GB RAM.

　　• Intel NetStructure PBX-IP Media Gateway, 8 Ports (Analog Model).

Figure 9.        Intel NetStructure PBX-IP Media Gateway front (above)



Figure 10.        Intel NetStructure PBX-IP Media Gateway rear view

The Intel NetStructure PBX-IP Media Gateway 10 was selected not for its compatibility with Nuance's software, but for its flexibility in connecting to various telephone lines. The Intel PBX-IP Media Gateway is a telephony gateway appliance that

connects to as many as eight analog phone lines through its digital telephony interface and connects to a LAN via a 10 BaseT or 100 BaseT Ethernet connector.

### 2. Software

Listed below are the software applications used to conduct this test:

- Microsoft's Windows XP

- Sun's Java 2 SDK 1.3.1_15

  - Sun's Java 2 SDK is a development environment for building applications, applets, and components using the Java programming language. This software is downloadable from Sun's website at http://java.sun.com/j2se/ 1.3/download.html.

- Nuance Voice Platform 3.0. with SP4 & Management Station

- Nuance Caller Authentication (NCA) 1.0 & Analysis Station

- Nuance Vocalizer 4.0

- Oracle's 9i Database

- Cygwin
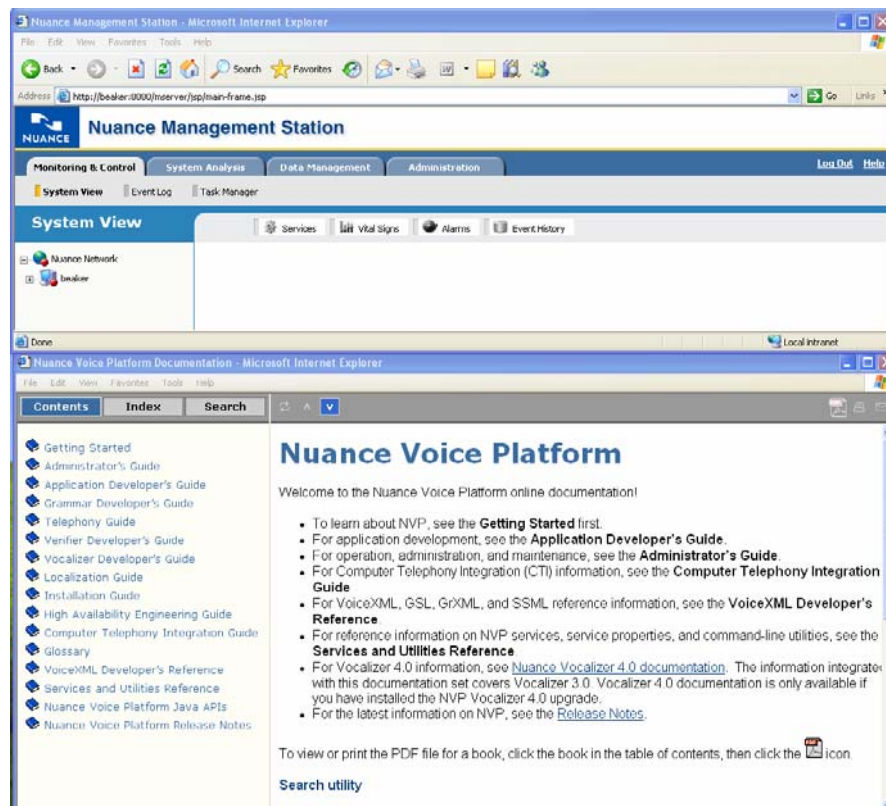


Figure 11.        Nuance Voice Platform 3.0. with SP4 & Management Station

## C. TEST ENVIRONMENT

The NPS Speaker Verification Test was conducted remotely. The NCA system was setup in the CENETIX Laboratory located in Root Hall Room 202 at NPS in Monterey, California. All calls made to the system were routed from the caller's selected communication medium (landline or cell phone) to the NCA system (located on the server) via six analog phone lines connected to the Intel PBX-IP Media Gateway. These six phone lines were requested through the Information Sciences department who, in turn, contacted the school's telecommunications department for the installation in the CENETIX lab. The coordinator was instructed to configure the system in such a way that only one phone number would be needed. If a person called the number and the first line was busy, the call manager (by Audix) would cycle the caller through the six lines until an unoccupied line was located. Since the calls did not take more than a couple of minutes each, there were not any complaints from the voice subjects regarding long wait times.

During the setup of the speaker verification test, special features of the NCA application were intentionally disabled in order to determine the raw estimates of the accuracy of the system without any fine-tuning. The two features that were disabled included: Variable Length Verification (VLV) and Online Adaptation [7].

- Variable Length Verification is a mechanism used by NCA for providing the most accurate results based on the fewest utterances. In the NPS Speaker Verification Test, this feature was intentionally disabled in order to collect more voice data for the offline impostor test.
- Online Adaptation is a feature that allows a system to adapt a stored voice model automatically during a verification session if it determines that the user is the true speaker. For the majority of calls, the system collected two utterances during the verification process.

## D. VOICE SUBJECTS

In order to conduct the test at NPS, a suitable number of voice subjects, approximately fifty, needed recruiting. Initially, the NPS Team thought that enough voice subjects could be recruited relying solely on the good will of Iraqi expatriates in

southern California (primarily San Diego, where a large community of Chaldean Iraqis live).  After several trips to contact potential voice subjects and phone calls to people connected to the Iraqi Chaldean community, it became obvious that good will alone was not going to suffice.  Many Chaldean Iraqis, being of Christian vice Muslim faith, did not feel a connection to their brethren back in Iraq.  Some had disowned their country completely and felt a deeper connection to the United States where they had made their recent fortunes in various business endeavors.

In fact, the only tie many of the potential subjects had with their native homeland was the fact that they speak the same dialect.  The question posed by most potential voice subjects was "What's in it for me?"  Because of this fact, additional funding was required from the project's financial sponsors.  These funds allowed for additional financial incentives to be offered to participants of the study.

On a chance meeting out in town, the author - Captain Pena - ran into a family he thought was Iraqi and struck up a conversation.  It turned out that the family was, in fact, Iraqi and worked for Defense Language Institute (DLI) in Monterey as Iraqi Arabic instructors.  After several follow-up meetings it was determined that the experiment could be conducted with the help of other DLI Arabic language instructors who were native Iraqi speakers.  After contacting the Provost of the Middle East School at DLI, it was determined that they had recently hired an influx of Iraqi Arabic instructors and that these faculty members would be willing to assist NPS in their project.

The compensation for the voice subjects would be based on their overtime pay and the amount of time spent conducting the verification and imposter trials.  The DLI instructors were accustomed to helping other government agencies by conducting experiments and by using their language talents for the benefit of scenarios used to train service personnel prior to deploying to the Middle East.  It was also an ideal fit because the age, education and experience level with modern information systems varied among this group and was representative of the education, age and experience level of the groups that would use this system in Iraq.

The goal for the NPS portion of the experiment was to reproduce more faithfully, the type of scenarios and environment that this system would encounter if deployed in

Iraq.  Therefore, although the voice subjects were given ample instruction in how to use the system and the type of line they should use to call the system (primarily wireless vice landlines) they were not coached during all portions of the experiment as was done under the Nuance test.  After the voice subjects were identified, two meetings were conducted with as many of the voice subjects as possible to discuss the key points of the experiments with them.  As can be expected to occur if the system is fielded in Iraq, not all of the voice subjects made it to the meetings due to conflicting schedules and other commitments.  In order to mitigate this problem, detailed instructions were handed out as part of their contract and other required paperwork.  (See Appendix D).  Listed on those instructions were contact numbers for the people conducting the experiment, to include a native Iraqi speaker in case any of the voice subjects encountered problems or had questions during their participation in the experiment.

Despite the steps taken to avoid confusion, a few of the voice subjects had difficulty fully understanding the test protocol:

- A handful of the voice subjects called in while a great deal of background noise was audible.

- Some voice subjects, in an attempt to isolate themselves from any background noise, called into the system from what appeared to be a bathroom or other room with a great deal of echo, even though it had been explained that this was not ideal for the system and would cause problems.

- A few voice subjects did not give a good voice enrollment because they cleared their throat while recording their voice, or counted from 1 to 10 instead of from 1 to 9, or their initial enrollment had a bad signal that did not allow for a quality enrollment.

- Other voice subjects were not consistent in speed, cadence, and volume throughout their enrollment and verifications (i.e. enrollment recorded at a very slow and hesitant pace and verifications done at a very fast, impatient speed and cadence and at a high and irritated volume).

All of these factors contributed to false rejects and possibly false accepts.  A great deal of these errors can be attributed to cultural and language differences.  Furthermore, it has been observed that Iraqis are eager to please their colleagues/bosses/clients etc.  As a result, it is difficult for them to admit or communicate that they do not understand what is being asked of them or that they are not capable of doing what is asked of them.  Whereas many westerners have no problem stating that they do not understand something

or that they cannot deliver what is asked of them, many Iraqis cannot bring themselves to admit this and instead, try to work through their difficulties, later upsetting their western counterparts/superiors/clients by not performing as expected.

As stated above, the most difficult error to deal with was the inability by some of the voice subjects to adhere to the agreed upon schedule. Some of the voice subjects decided to finish conducting the verification calls during the imposter trials. This caused the false-acceptance report to appear much worse than it actually was and required a great deal of time for the review of each call to determine which ones were true imposters and which were simply late callers. In hindsight, it would be best to have a bigger break between verification and imposter trials or even to arrange a separate group of imposters to conduct the calls to reduce the chance of errors due to overlap.

## E. TEST SCHEDULE

In order to isolate the verifications from the impostor trials, the voice subjects were instructed to call during the first three weeks of the experiment and make imposter trials during the last week of the experiment. Between the first and third weeks of the experiment, a break was scheduled during which no one called into the system in order to give the subject's voice a chance to change through the course of the experiment. This decision tested the system more fully by proving its ability to deal with natural variations in a subject's voice due to time, illness (stuffy nose and so on), and other variations that occur naturally throughout the day (i.e. the difference in a subject's voice when he/she first wakes up compared to after a full day of speaking in a classroom).

## F. TEST PROTOCOL

The test protocol for the speaker verification test consisted of four steps. In step one liaison was made with DLI requesting test subjects to volunteer their time in exchange for financial compensation to participate in this experiment. The initial meeting provided the students' liaison, Mr. Detlev Kesten, with a general overview of the Independent NPS Speaker Verification Test, to include a demonstration of a verification call made in Arabic. As part of the NPS/DOD regulations for the use of human subjects,

the NPS research team obtained permission from the NPS Human Resource Board prior to conducting any testing; the submission packet is included as Appendix B of this document.

In step two, several meetings were held to give the information on the conduct of the testing, to include sample call dialogues of the speaker enrollment and speaker verification process, and applicable participation consent forms. Once all the consent forms and contracts were signed and instruction sheets were handed out (examples in Appendix C, D, and E respectively), the participants were divided into two groups, cell-phone users and landline users. This was done on a 4 to 1 basis in order to match the current situation in Iraq where, due to limited infrastructure, there are more cell phone users than landline users. Both groups were asked to dial a given telephone number to enroll and to verify their voice biometric. Participants were given the opportunity to try the system out before the test officially started in order to limit confusion once the test actually began.

In step three, participants were asked to enroll once and then verify ten times during the first week of the test (07-13 May 07) and to verify again ten times during the second week of the test (21-27 May 07). As stated before, the participants were given a week off (14 -20 May) to allow their voices to change. This would provide for greater test accuracy and it also allowed for built in flexibility should anything need adjustment or further explanation. During the enrollment process, participants were asked to register with the system using a unique 8-digit identification that was assigned to them at the onset. Participants were then asked to count from one to nine three separate times. All of the instructions were given in Arabic and all participants were native Iraqi Arabic speakers. During the enrollment, the three instances of voice samples were used for generating a unique model of the participant's voice pattern. During the verification process, the participants accessed their accounts with the unique ID and then were asked to count from one to nine twice.

In step four (28 May – 03 June 07) each participant was given a list of twenty-five account numbers into which they were to try and gain access. Some effort was made to try to match female callers with the accounts of other females, but both female and male

33

callers attacked all accounts. There was also a group of five individuals, dubbed advanced imposters that were allowed to listen to the enrollments and then attempted to gain access to those accounts. This was done to replicate the scenario where the imposter knows the voice and account number of a particular subject and is trying to mimic their voice, cadence, and speed. The last step of the experiment consisted of analyzing the data collected and reporting the results to all concerned parties.

## G.    TEST ANALYSIS

Upon completion of the test at NPS, the students were left with the raw data collected by the Nuance Caller Authentication (NCA) system. NCA also came with an analyzer tool that allows one to see the basics of the experiment, such as total calls made, successful enrollments, failed enrollments, successful verifications, failed verifications and so on. However, upon first glance at the reports generated by the system, it is not possible to glean which calls were truly false rejects and false accepts. In order to get a true picture of the results, Dr. Prieto of Nuance generated a script. This script identified the calls that were rejected during the verification phase or the calls that were accepted during the imposter trials that gave them their potential false rejects or accepts. However, these initial results were very misleading. It was still necessary to listen to each call to determine if the reason the calls were rejected had something to do with a bad phone line, improper technique on the part of the voice subject, or other factors.

Further, it had to be determined whether any of the voice subjects made verifications to their own accounts during the imposter trials. It was also important to identify if there were any other factors that would make the system fail and thereby become a critical vulnerability, such as speaking very fast or slow or having some noise in the background.

The script given to the students by Dr. Prieto was a Linux based script run with Cygwin. Once a time period was identified, the script could identify which callers were rejected during the verification phase and which callers were accepted during the impostor trials. The result was two Excel files, one each for potential false accepts and rejects. The files listed the calls that needed further study and had hyperlinks to listen to the voice file created for that particular call. This made it much easier to run through the

34

hundreds of calls without having to search through several directories and use separate programs for audio and the reading of the database in order to glean which calls were true verifications and which were not.

After listening to the false rejects, several calls were disqualified because of problems with the quality of the phone line during a particular call or because of an exaggerated deviation to the prescribed volume, speed or cadence of the utterance. For example, several calls had a great deal of noise in the background, while others had beeps from another incoming call during their utterance. Still others, perhaps out of nervousness, yelled their utterance much slower and louder than their enrollment and in direct disregard to the instructions given to them. These particular situations were unique and it was determined that they should not be counted against the system's accuracy.

Determining which false imposters to disqualify was a lot more difficult. It had to be based on human judgment and anecdotal data from the experiment. For example, a few days into the imposter trials a couple of voice subjects called and asked if they could begin their imposter calls. This led to the discovery that some of the voice subjects were not following the prescribed schedule despite clear written instructions, verbal explanations in English and Arabic, and several emails detailing the schedule and reminding the voice subjects what they should be doing that week. Upon reviewing calls, it was realized that those questionable callers had in fact made a great deal of their calls during the imposter phase that skewed their results considerably. Additionally, the subjects had been instructed that any caller that was able to gain access to any of the thirty accounts during the impostor trial should attempt to access that account again. They were instructed to do this in order to determine whether the access was a one-time fluke or, in fact, something they could achieve every time they called back.

After the calls made in error were discarded, the duplicate imposter calls were thrown out in order to get a true picture of the results. The argument was that duplicate calls should not be counted because if they were, a user that gained access into someone else's account could call back hundreds of times and completely skew the results. In fact, one caller did something similar. After he gained access the first time, he took it upon himself to call 20 more times until the system rejected him again. All of his duplicate

calls were deleted as well. After the data was cleaned and only legitimate false accepts and rejects remained on the Excel file, an additional script provided by Dr. Prieto was run in order to give the ROC curve. Table 2 is a spreadsheet that describes how the final numbers were determined. The first column delineates the area of concern. The subsequent columns enumerate the findings of the final results of the Nuance test (Nuance Analysis), the original results of the NPS Test (NPS Analysis) and the final results of the NPS test (NPS Analysis Excluding Outliers). Enrollments refer to the total number of voice enrollments recorded by a test-subject for an individual account. The "Number of Calls" refers to the total number of calls received by the system. Valid Verification Attempts refers to the total number of calls that were intended by the user to access his or her account. False rejects is the number of those who tried to gain access to their account, but were denied access. Imposter Trials are the number of calls made by those trying to gain access to the incorrect account with access to that account number. The number of those calls that were successful is the "False Acceptance." The "Accuracy Analysis refers to the calculations of the system accuracy given the results of each test. The confidence interval refers to the ability to achieve those same results given similar testing environments. The False Acceptance and False Rejection Rates (in percentages in the row for False Acceptances and False Rejections), as well as, the overall system accuracy and the confidence interval of that accuracy were made using formulas described in Chapter II.

| Discussion | Nuance Analysis | NPS Analysis | NPS Analysis Excluding Outliers |
|---|---|---|---|
| Enrollments | 239 | 44 | 41<br>Note: three poor quality voice enrollments were discarded |
| Number of Calls | 14,130 | 2,658 | 2,559<br>Note: 99 calls were discarded |
| Valid verification attempts | 2355 | 1324 | 1377<br>Note: 98 calls made during imposter trials were meant to be verifications. 45 calls were discarded due to quality or other concerns |
| False Rejects | 129 (5.48 %) | 57 (4.3 %) | 11 (0.8 %) |
| Imposter Trials | 11,775<br>Note: Nuance's imposter trials were simulated offline attempts using utterances collected during verification trials. | 1334 | 1182<br>Note: 98 calls made during imposter trials were meant to be verifications. 54 other calls discarded due to quality or other concerns. |
| False Acceptance | 236 (2.0 %) | 262 (19.6%)<br>. | 59 (4.9 %)<br>Note: 98 calls made during imposter trials were meant to be verifications. 54 calls discarded due to quality or other concerns. 51 duplicate False Accepts were also discarded. |
| Accuracy Analysis | FRR: 5.48 %<br>FAR: 2.0 %<br>Accuracy: 97.41 % | FRR: 4.3%<br>FAR: 19.6%<br>Accuracy: 88.00 % | FRR: 0.8 %<br>FAR: 4.9 %<br>Accuracy: 97.26 % |
| Confidence Interval | 0.54%<br>Accuracy:<br>97.41 % ± 0.54 | 1.17%<br>Accuracy:<br>88.00 % ± 1.17 | 0.62%<br>Accuracy:<br>97.26 % ± 0.62 |

Table 3.    NPS Speaker Verification Test Analysis Comparison

Specifically, the following calls were discarded or migrated to their correct phase:

Three Accounts Deleted:

- **00606531** discarded due to poor quality of enrollment and verifications. Enrollment recorded very slow and low while verifications attempted in a loud, impatient voice and inconsistent speed and cadence (11 verifications deleted).

- **12433668** discarded due to echo in verification as well as enrollment. Also clears throat and counts to ten vice nine during enrollment (3 verifications and 17 imposter trials deleted).

- **13181752** discarded due to a great deal of background noise in enrollment and caller counts to ten vice nine (11 verification calls and 6 impostor trials deleted).

Verification calls deleted due to individual problems with the call:

- 1 call from acct. # 00680310 discarded due to high volume and incoming call during verification.

- 15 calls from acct. # 12135912 discarded due to too much echo.

- 4 calls from acct. # 20350272 discarded due to too much echo.

Imposter calls moved to verification phase because the callers violated the schedule and called their own accounts during the imposter trials:

- 15 calls from acct. # 11687972

- 25 calls from acct # 13192682

- 4 calls from acct # 13037119

- 34 calls from acct # 22651638

- 12 calls from acct # 31198392

- 4 calls from acct # 32368732

- 2 calls from acct # 33284776

- 2 calls from acct # 33692974

Other False Acceptance calls deleted:

- 17 calls from acct # 12433668 due to account deleted because of bad enrollment

- 6 calls from acct # 13181752 due to account deleted because of bad enrollment

## H. ESTIMATES OF CONFIDENCE INTERVALS FOR THE NUANCE IRAQI ARABIC VOICE VERIFICATION TEST FOR PHASE 2 C

The Phase 1C test had 239 speakers. The total number of voice verification attempts was 2355. The total number of imposter attempts was 11775. The NPS test had 44 speakers with 1324 voice verification attempts. The NPS test, excluding outliers, had 41 voice subjects and 1377 voice verification attempts. The confidence interval computed using Normal Approximation for the various test data sets are given in the last row of Table 2 above.

## I. COMPARISON WITH PREVIOUS SPEAKER VERIFICATION TESTS USING NUANCE'S TECHNOLOGY

### 1. Nuance

As seen in the table above, Nuance's test consisted of 239 native Iraqi Arabic speakers that were residing in Jordan during the experiment. Those voice subjects made 2,355 live calls to the system under very controlled conditions. In addition, the imposter trials were made offline (not live) using voice utterances from the verification trials to try to break into other accounts. Unlike the test at NPS, the majority of the callers in Jordan was brought into a call center where a caller could be coached or get help from test proctors. While this made for a smooth experiment and less user error, this is not how the system would normally be used in an operation with ministers of the Iraqi government. The impostor trials also did not faithfully replicate some of the craftiness of which humans are capable, as did the advanced impostor trials done at NPS. In their defense, Nuance was not allowed to use the tuning mechanisms that would normally be used in a live system that would continuously improve the reliability and accuracy of the system as it learns the account holder's voice. A full explanation of Nuance's experiment and performance report can be found in Appendix A.

## 2. Past results Compared to NPS Results

As shown in the table above and the graph on the next page, the NPS test did not replicate the same results as the Nuance test with the Jordanian voice subjects nor the past phases (Phase 1A and 1B) of the IEVAP project. However, considering that this test was done with a new language module developed by Nuance specifically for this experiment, it performed well. Despite the different methodologies employed between the NPS and Nuance test a comparison the ROC curves does promote a level of confidence with respect to the overall system accuracy.
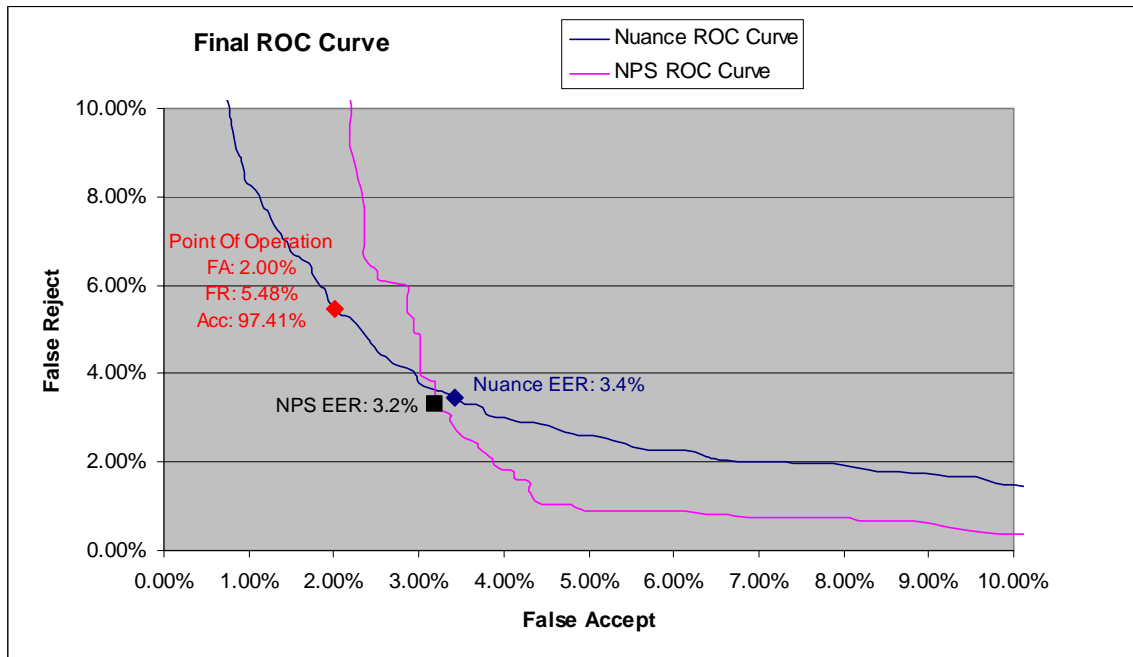


Figure 12.        Comparison of Nuance and NPS test for Iraqi Arabic (Phase 1C)

ROC Curve

NPS Data-set: 3.0% EER
Data-set consist of :
68 enrollments, 411 speaker trials
4,300 impostor trials

Nuance Comparison Data-set: 3.4% EER
Data-set consist of :
139 enrollments, 565 true speaker trials
80,730 impostor trials
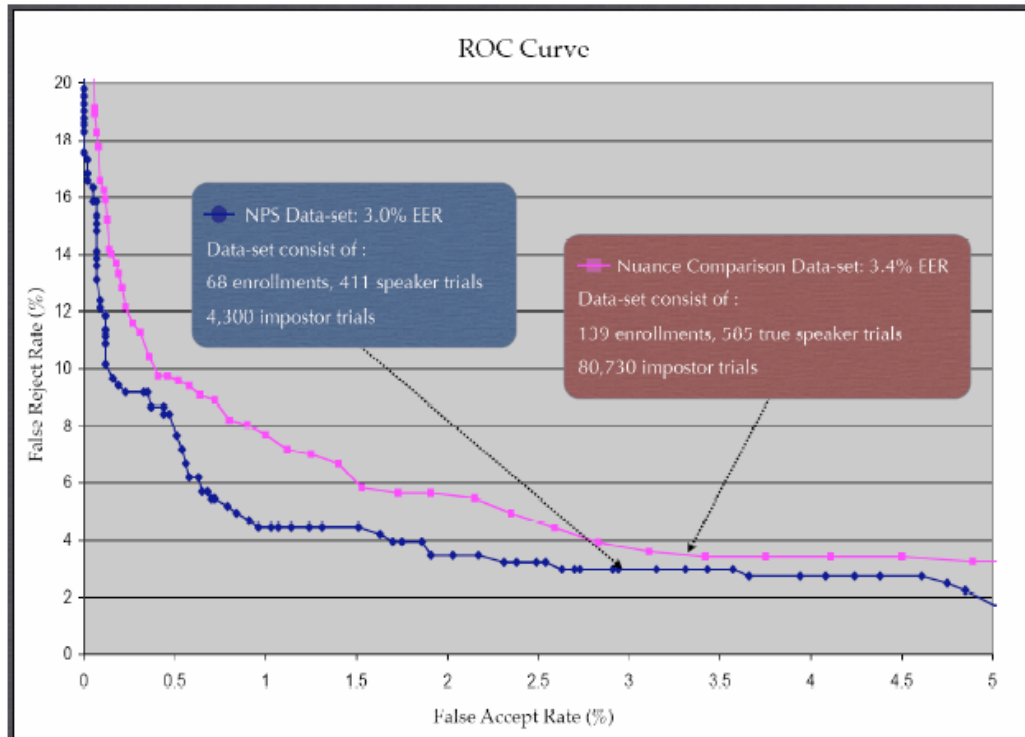
Figure 13.        Comparison of Nuance and NPS test in English (Phase 1B) [From 7]

## J.        TEST LIMITATIONS AND ASSUMPTIONS

### 1.        Test Limitations

The largest limitations of this research effort were time and money.  With more time, a great deal more voice subjects could have been recruited, allowing for a fuller test of the system.  In order to make up for the time and financial constraints, the voice subjects were requested to make more test calls per person.  After discussing sample size concerns with a statistics professor (Lieutenant Colonel Lee Ewing), it was learned that in order for the experiment to meet the ideal sample size at least 40 voice subjects would be needed.  Furthermore, it was important to have the total number of voice subjects make at least 1,024 calls in all during each phase.  The NPS experiment exceeded both of these criteria and the system was ultimately tested more severely than a live system would be.  This stems in part because the proportion of imposters to true callers would rarely be as significant as this experiment that had nearly a one-to-one proportion of valid verifiers to imposters.  In addition, it is rare to have imposters with access to the all of the voice files

41

as the advanced impostors did.  This allowed the advanced impostors to pick a voice that was similar to theirs and try to mimic it in order to break into the system.  It should be noted that all false accepts happened during random imposter trials and not during these advanced impostor trials.  The severity of the imposter trials made up for the lack of voice subjects and fairly tested the reliability of the system.

### 2. Assumptions

Since all 59 imposters were able to access the account they breached more than once, it was assumed that access to an account, for the most part, meant full access as often as the imposter wanted it.

## K. PHASE 1C SUMMARY

In Phase 1C of this project, NPS successfully conducted a speaker verification test to assess Nuance's speaker verification technology based on the performance measures of FRR and FAR.  During the test, NPS did not impose any restrictions on the environment from which the calls originated.  Also, while the Nuance ROC analysis yields an equal error rate of 3.4 % (FRR based on 2,355  trials, FAR based on 11,775 trials) and a system accuracy of 96.22 %, the NPS analysis yields a FRR of 0.8 % and a FAR of 4.9 % (based on 1377 verification attempts) and a system accuracy of 97.26 %.  The ROC analysis equal error estimates of the NPS test are in the same range as the average estimates of the equal error rate by Nuance based on other similar datasets.  This validates the NPS test in spite of the smaller number of enrollments and speaker verification attempts.

# V. CONCEPT OF OPERATIONS

## A. PHASE 1C OVERVIEW

Initially the purpose of Phase 1C was to test the Iraqi-Arabic Speaker-Verification Application developed by Nuance and to further the work on the Baghdad Central Correctional Facility (BCCF) as described by Captain Sam Lee, USMC, in Phase 1A. After the project began however, representatives from OSD, the sponsor of this project, suggested the direction shift to evaluate this application as a means to further the banking system in Iraq. This philosophy is in keeping with both the National Military Strategy and the Strategy for Victory in Iraq.

The later of the two documents has three tracks: "political, security, and economic" [20]. The economic track has six core assumptions, the fourth of which is: "economic change in Iraq will be steady but gradual given a generation of neglect, corrosive misrule, and central planning that stifled entrepreneurship and initiative"[20]. The problems of this misrule have led to a culture of corruption. As stated in Chapter I, billions of dollars either have been lost to mismanagement, theft, or have simply gone unaccounted for. One of the continued challenges in Iraq is "Creating a payment system and a banking infrastructure that are responsive to the needs of the domestic and international communities, and that allow transactions involving possible money laundering, terrorist financing and other financial crimes to be detected" [20]. That being said, although this system developed could still be used on the menu driven system developed for the BCCF, the focus would now be on how to use the Nuance system with regards to Iraqi Banking and its role for victory in Iraq.

## B. THE ROAD AHEAD

Chapter IV discussed in detail the findings from Nuance's test and the NPS independent test of the Iraqi-Arabic Speaker Verification Package. These findings were such that it is recommended that Nuance's Iraqi Arabic Speaker Verification System be used as the front door for the new era of banking in Iraq. There are currently four options available for Phase II:

43

| Option 1: Voice and or Touch Pin<br>Leverages existing Nuance software Voice Authentication Engine; Custom development of Front End (Robust) | ROM Cost<br>$ 0.8 Million | Time<br>6 months |
|---|---|---|
| Option 2: Touch Pin<br>Leverages existing Nuance software Voice Authentication Engine; Custom development of Front End (Limited) | NA | NA |
| Option 3: Voice and or Touch Pin<br>Leverages Next Generation Nuance software Voice Authentication Engine; Custom development of Front End (Robust) | $ 1.4 Million | 12 months<br>Upon Release of New Software |
| Option 1 System Upgrade<br>Conversion from existing Nuance software to Next generations software | $ 1.2 Million | 12 months<br>Upon Release of New Software |

Table 4.    Phase 2: Application Development for Iraqi Arabic only [After 21]

Option 1 completes the existing entry control point with existing software and allows for a robust front end. The advantage of this option is that for a fairly low cost, a user can have a working system (front end) in a short amount of time. Option 2 again uses existing software and provides for a limited front end. This can be done at almost no cost and it merely adds a pin to what has already been done. Option 3 is the development of a robust front end using the next generation of Nuance software. The advantage of this option is that the purchaser is not buying obsolescence; he or she is using the latest technology for the implementation of the banking system. The drawback to this option is that it will take twice as long to create a fully functional system versus the first option. The final option is to purchase option 1 now and implement in six months. As funding becomes available, the upgrade in software can be transitioned from old to new at user direction. The difference in cost is nominal given that there is a waiting period that could delay the start of the project. Given the current political situation in Iraq, the authors are recommending Option 1 on the assumption that there is a bank to which this system can be attached. In addition to using this system in Iraq, there are also deployment considerations worth exploring with respect to other Middle East countries such as Afghanistan.

As of CY2005, there were 1.4 million cell-phone users, 280,000 landlines and 30,000 Internet users within Afghanistan [22]. "By end-2010: a national telecommunications network will be put in place so that more than 80% of Afghans will have access to affordable telecommunications, and more than US $100 million per year are generated in public revenues" [23]. This means that a majority of the people in Afghanistan who have access to telecommunication systems are telephone users, making the country ripe for voice-based technology as well. Below is a list of the same options given the development of a front end using three languages (Iraqi Arabic, Dari, and Pashto) indigenous to the region:

| Option 1: Voice and or Touch Pin<br>Leverages existing Nuance software Voice Authentication Engine; Custom development of Front End (Robust) | ROM Cost<br>$ 2.3 Million | Time<br>12 months |
|---|---|---|
| Option 2: Touch Pin<br>Leverages existing Nuance software Voice Authentication Engine; Custom development of Front End (Limited) | $ 1.6 Million | 6 months |
| Option 3: Voice and or Touch Pin<br>Leverages Next Generation Nuance software Voice Authentication Engine; Custom development of Front End (Robust) | $ 2.9 Million | 12 months Upon Release of New Software |
| Option 1 System Upgrade<br>Conversion from existing Nuance software to Next generations software | $ 1.6 Million | 12 months Upon Release of New Software |

Table 5.     Phase 2: Application Development for Iraqi Arabic, Dari and Pashto Languages [After 21]

Much like the application being developed solely for use in Iraq, time is still a factor for implementing this system. The advantage of not waiting for the system upgrades is a cost savings of almost 1 million dollars. As discussed in previous chapters, Voice or Speaker Verification can offer a number of options when it comes to security

and banking services.  This concept of operations will specifically discuss how this capability can be implemented as an entry control point for an Iraqi banking system.

## C.    CONCEPT OF OPERATIONS

The two most basic questions that need to be answered are:  1) Who will use the system?  This system will first be implemented within the government itself, including the payment of all employees, both government civilians and military members, and transferring of money for the payment of outside contractors; 2) How will the system work?  In essence, this system would be the front door to a telephonic banking system. The user would simply call a number to access his or her account.  At the door created by Nuance© the user would first be authenticated and once inside, the user could move around and manage their account.  Account management would include the ability to transfer money to other accounts in order to pay bills, check account balances and transactions, and verify receipt of payroll checks.

In the case of government accounts at the uppermost level, money transfers would have to be made via computer.  Once the transfers were made to individual departments, such as the Department of Defense or Department of Energy, the voice authentication system could be used to further distribute funds.  In the case of military personnel, police officers and government employees, salary payments would be based on how much time an individual worked and would be paid directly into his or her account from a central facility like the one the U.S. military uses in Kansas City.

It is important that there not be any roadblocks to paying employees.  For every person that has to verify a particular transaction, the process of payroll is slowed, halted or possibly corrupted.  As will be discussed later in Chapter Six, these employees are the sales force for this new technology.  If they disapprove of the system or it creates a situation in which they are not paid regularly, the system will ultimately fail.  On the other hand, because of the environment in which this system will exist, there must be sufficient checks and balances to ensure that each transaction between departments and contractors is accounted for and verified.  Each level of transaction will require different security measures.

For paying government workers, to include police and military, the system will have to be set in order to allow for the greatest amount of usability, meaning that the False Reject Rate will be lowered.  As discussed in Chapter II, this means that the False Acceptance Rate will be increased, meaning that the chances of someone gaining access to a personal account will be greater.  The system would still require that the unauthorized user know the account number, but the chance of criminals accessing accounts will indeed be greater.  Because these accounts are personal accounts, the total dollar amount affected will be lower and therefore the risk for loss is worth granting greater accessibility.

For those accounts that are used to transfer money from within the federal government to an outside contractor, the security level will have to be much higher.  At this level, usability is less important than keeping intruders out of the system.  This increased security will require two things - greater security on initial entry into the system and knowledge verification.  In order to increase security on initial entry into the system, accounts that deal with large amounts of money will have to have a very low rate of False Accepts.  This means that the False Acceptance Rate would set to be very low.  Conversely, the False Rejection Rate will be much greater.  This, of course, will lead to a greater False Reject Rate, but the risk of loss in this case is much greater than in personal accounts.  Therefore, in addition to the account number and voice print, either a pin or another form of authentication known as knowledge verification will need to be used.

Knowledge verification is the process of extracting pertinent information from the account holder, such as verifying a pin number or a mother's maiden name.  Further, because the system is being implemented in a society where government officials are kidnapped and killed everyday, duress codes will have to be added.  Although there has been headway made in the area of detecting porosity in voices in order to detect duress, it is not currently accurate enough to be used as an alert.  To combat this potential problem, the system can be designed to allow a unique pin to be used as a duress code much like those that are used in personal alarm systems. This provides the perpetrators the illusion that they have made successful entry into the system, but it will also alert the bank and proper authorities that the user is under duress and in need of assistance.

## D.    INITIAL ENROLLMENT

For this application, the first task that would have to be done is the user would have to be identified to be the person that they are claiming to be.  This is important because if the person is falsely identified from the onset all future authentications will be fraudulent.  Once properly identified the person would be given an account number; this number would coincide with their newly created bank account.  In order to gain access to the account, the user would have to call in.  For the first time using the system, the user would be asked to provide his or her account number.  When this information is queried against the database to determine if a voice print is on file -- and none is found -- the system will prompt the user to create one.  This is the same as the enrollment process discussed in Chapter IV for the initial tests of the system.  Depending on the security requirement, this might need to be done immediately after the account is made in the presence of security officials or administrators for the system.  This will ensure that there is not a period time where the account is vulnerable to an imposter with a list of account numbers calling in hopes of getting his or her voice imprint on an account.  In addition to the initial voice print, this would be a good time to set up a secondary verification, such as a pin number.  As mentioned in Chapter II, the voice recording that is created by the system before it extracts the needed information to create a voice template for the user can be used by other programs with other algorithms for the purposes of voice identification.

## E.    VERIFICATION

The second time the user calls in to the system the user will be asked for the account number.  Once the account number is verified, the user will then be asked to count from 1 to 9 in Iraqi-Arabic.  Once the user has been authenticated, the user will be transferred to the banking system.  If the user is having difficulty, the system should then turn the user over to customer service for further assistance.  Customer assistance should be trained on how to access the system in order to listen to the voice print.  If the initial enrollment is not clear, the user should be instructed to go to his or her bank in order to re-enroll into the system.

## F. PLANNING FOR THE SYSTEM

If any of the options that Nuance has recommended are acquired, Nuance recommends the following steps for the planning and provisioning of the system [24]:

- Budgetary Sizing- a rough estimate made during the presales activities
- Engagement Sizing- an adjusted estimate based upon the requirements analysis
- Final Sizing- a detailed, accurate provisioning based upon pilot data.

These steps serve as an iterative approach to coming up with the requirements for the system once it is in place. During each of these steps, the "Major Planning Tasks" must be made. Those steps are [24]:

- Analyze the Telephony Requirements
- Analyze the Application/system Requirements
- Determine the Network Topology
- Provision Clusters
- Define the Management Station User Roles.

For the purposes of this thesis, each major planning task will be discussed in order to seed the discussion for a future system implementation.

### 1. Telephony Requirements:

The first step is to determine the In-bound Telephony Channel Requirements. To do this the following must first be identified:

- Peak Call Volume V (calls per second)
- The average call duration t (seconds)
- The allowed blocking probably

The first thing that must be calculated is the traffic on the system. This is known as "Busy Hour Traffic (BHT) (in Erlangs) is the number of hours of call traffic there are during the busiest hour of operation of a telephone system" [25]. The following formula is used:

$$BHT = V * t$$

That information is then entered into an Erlang-B calculator, like the one found at www.erlang.com. "The Erlang-B formula is a model used by telephone system designers

to estimate the number of lines required for PSTN connections (CO trunks) or private wire connections" [25]. Nuance refers to these lines as channels. For example if V = 1 calls/sec and t = 120 sec/call, A = 120 erlangs. Then, using a blocking probability of .03, which means that 3 out of 100 calls will be blocked, those numbers are plugged into the erlang calculator resulting in the need of 130 lines or channels.

The next step is to determine the transfer channel requirements. There are three types of transfers - blind, bridging and Two-B Channel Transfer (TBCT) [24]. A blind transfer occurs when the moment a user is identified they are transferred directly to the banking system. Therefore, no additional channels are needed. A bridge transfer connects an in-bound and out-bound line for the duration of the call. This requires double the number of channels calculated for using the erlang calculator, which would be 260 channels. For the TBCT, the call is dropped once the connection is made. Because the Nuance system will only act as the front door to this system and the only way into the Iraqi Banking system should be through the front door, it is recommended that a blind transfer should be used initially.

Once this is done, the user must determine what telephony system will be used with this application and provisioning must be made. A Publicly Switched Telephony Network (PSTN) allows for up to 4 T1s per telephony card. A T1 line allows 23 channels. Voice over IP (VoIP) uses Session Initiated Protocol (SIP), which allows for 69 channels. Using the example above, six T1s would have to be used and therefore 2 telephony cards would have to be used or 2 VoIP hosts would have to be used.

### 2. Analyze Recognition Requirements

In order to analyze the recognition requirement, one must determine both the recognition and the grammar load. The recognition load is measured in recognition units (RUs). "1 RU is the amount of recognition power required to understand a continuous sequence of digits in real time with a 1% error rate" [24]. The RU depends on three factors: type and speed of the CPU; overall hardware configuration; and version of Nuance software installed. A grammar load is measured in Load Units (LUs). "1 LU is the load of a grammar that can be recognized in one CPU that has a recognition power of 1 RU" [24]. LU is a function of the complexity of the grammar and the recognition

50

parameters used. For example, a 16-digit string in Swedish takes 1.92 LUs and that same string in French Canadian takes 1.24 LUs. The application requirements for the system such as the text to speech requirements (TTS) will determine the requirements for host memory. This will determine what type of CPU is needed to run the application. In order to fully calculate this number, a Nuance Dimensioner must be used.

### 3. Determine Network Topology

"Each Nuance Voice Platform (NVP) is a self-contained entity, complete in itself, compromising all the elements needed to deploy service, including application servers, database servers, and cluster hosts." [24] There must be at least two clusters per node. This is required such that if one host is taken off line for any reason, another host is up and running. This allows for maximum uptime. This also means that each host must be identical to the other. Each host can handle a maximum of 24 T1s (552 channels) per cluster. Using the example above, 130 channels, and two clusters would be required.

### 4. Provision Clusters

In order to determine the provisioning of an NVP cluster the Nuance recommends using the following guidelines [24]:

- Management Stations – 1 per Cluster
- Browser Hosts
    - NMS: 1 host per 92 Channels (4 T1s)
    - SIP: 1 host per 69 Channels
- Recognition Hosts
    - Number of hosts per cluster = (Application RUs per Cluster/CPU RUs) + 1
    - 2 hosts must be configured as Resource Hosts
- Audio Output Hosts
    - Number of hosts per cluster =

        (Incoming Channels per cluster/Channels per host) + 1

**5. Define the Management Station User Roles**

The final major planning task is defining the Management Station User Roles. Each NVP Cluster will require the following personnel [24]:

- System administrators that configures the host and has privileges to access all other systems.

- System operators that control hosts and services, manage data, and generate and view reports.

- Application Tuners and Dialog designers who view and schedule reports, browse call logs and listen to calls

- Application Developers that view event logs and scheduled reports

- Business Users that view schedule reports.

The number of personnel required for each of these positions will vary based on the size of the system that is being implemented.

Once all of these tasks have been completed, budgetary sizing is complete and the sizing process can continue.

# VI.  IMPLEMENTATION

## A.  OVERVIEW

There has been a great debate over Operation Iraqi Freedom and whether or not it was prudent for the United States to become involved in Iraq.  That debate aside, however, the fact remains that the U.S. did get involved, Saddam Hussein was overthrown and an entire country now needs rebuilding from the ground up.  What has also become apparent is that there is a lack of financial accountability regarding the money the Iraqi Government was supposed to use in an effort to rebuild their country. According to a recent report on corruption, billions of dollars earmarked for reconstruction are unaccounted for at this time [1].  The problem is so severe and so widespread in the upper levels of government that the current investigations have been stopped by an antiquated law and cannot resume until receiving the approval of the Prime Minister himself.

The problem with these investigations is that they involve "eight ministers and 40 directors general who are accused of mismanaging eight billion dollars" [26].  Prime Minister Nouri al-Maliki has stated "We suffer in terms of security and administrative corruption"[26].  Although the technology available through Nuance in terms of Voice Authentication has security implications, it also has the ability to provide the Government with accountability for its financial transactions by adding the feature of "non-repudiation," as mentioned in Chapter IV.  This change, although technical, must take hold with the people of Iraq or the change will not be a lasting one.

An expert on the subject of creating changes that will endure, Dr. Senge claims that several disciplines need to be mastered in order for an organization to be able to conduct meaningful and lasting change; in other words, to become a learning organization.  The most important of these five disciplines is developing systems thinking.  "Systems thinking" is the key to breaking away from the status quo and creating lasting change.  Senge states that systems thinking "is a conceptual framework, a body of knowledge and tools that has developed…to make the full patterns clearer, and to help us see how to change them effectively" [27].   The country of Iraq is in desperate

need of lasting change. In order to break free from previous mental models, the Iraqi people need to have a "Shift of Mind." Senge states "the unhealthiness of the world is in direct proportion to our inability to see it as a whole" [27]. Perhaps this idea is no more obvious than in the present country of Iraq. Nevertheless, in order to bring about successful change, first a diagnosis of the problem must be made.

**B.     DIAGNOSIS**

The first step in diagnosing a problem and coming up with a solution is to select a model that is appropriate to the particular problem. The Congruence Model, developed by David Nadler and Michael Tushman most directly matches the problem that IEVAP is trying to solve with its banking application in Iraq [28]. The next part of this chapter will be dedicated to discussing the model and how it applies to the Iraqi banking situation.
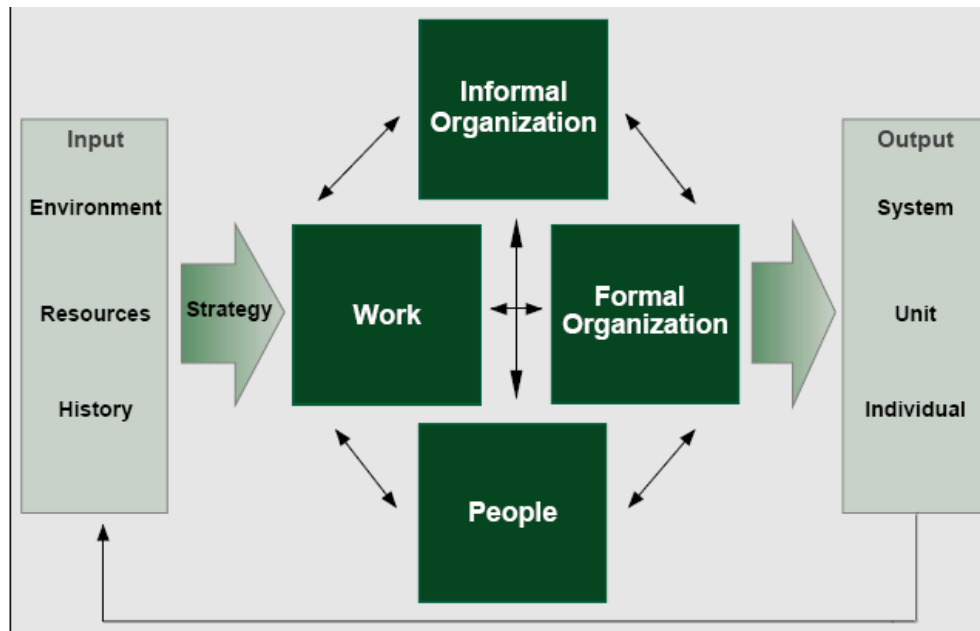
**C.     THE CONGRUENCE MODEL**



Figure 14.     The Congruence Model [From 28]

### 1.     Input

The first part of the congruence model is the input.  The input consists of three elements:  the environment, resources, and history.

- **Environment**

Within the congruence model, the environment "includes people, other organizations, social and economic forces, and legal constraints" [28].  To say the least, the environment in Iraq is hostile and presents a unique set of problems.  To complicate matters, the Iraqi environment includes an element of time restriction as well.  At the time of writing this thesis (September 2007) the U.S. President's approval ratings are at an all time low and the percentage of Americans who support the war is becoming less with each month the war continues.  In short, the American people are demanding a solution to the situation in Iraq.

Even more importantly, in the country of Iraq itself a gross number of people are dying daily, billions of dollars are unaccounted for and there exists a culture of corruption.  Further, there presently exists no banking system.  All of these factors present an unusually difficult environment to try and negotiate.

- **Resources**

In this model, resources include "the full range of accessible assets—employees, technology, capital, and information" [28].  As of this moment the people of Iraq have two resources that are crucial to the success of this project - they have money and they have access to telephones.  Having these resources allows the opportunity to create a telephonic banking system that will quickly become another important resource for the people of Iraq.

- **History**

Nadler and Tushman state that "(t)here is considerable evidence that the way an organization functions today is greatly influenced by landmark events that occurred in its past."  In this case, Iraq was a country that lived under the Iron Fist of Saddam Hussein al Tikriti for almost forty years.  Fortunately, his corrupt regime no longer has control and the new history of Iraq is now being written.  Unfortunately, however, Saddam's culture

of corruption persists to this day. This new banking system aims to hold accountable those who want to live in the past, continuing to support a culture based on cruelty and corruption.

As the model suggests, however, the history of Iraq continues to affect its people. For example, when one of the coauthors of this thesis was in Iraq for Operation Iraqi Freedom II, then Captain Withee was in charge of an Ammunition Supply Point (ASP). This ASP was supposed to have ninety Iraqi Soldiers from the Iraqi Civil Defense Corps (ICDC) working at it. These ninety ICDC were broken down into two platoons of forty-five Iraqi soldiers a piece. The Iraqis in these platoons were supposed to come to work every other day. Of those forty-five ICDC only 10 to 14 came to work everyday. The problem was that their battalion commander offered his troops a bribe. In return for giving the battalion commander half of their paycheck, which he was responsible for paying, he allowed the soldier not to come to work. According to Captain Fariz, the Company Commander of this group, this same type of corruption was rampant throughout the Iraqi Army.

In order to solve this problem, the U.S. government decided to consolidate payments for the Iraqi soldiers. Thus, in order to be paid, all the soldiers had to travel to a central location. This became a "fix that failed" because the insurgents used these "pay days" as an opportunity to attack soldiers who were pooled in large groups. In addition, each soldier could be missing for days at a time every payday, as they had to travel to the payment disbursement location and then deliver the money to their families in various locations throughout the country.

### 2. Strategy

The strategy within the congruence model is defined as "a set of decisions about how to configure its resources in response to the demands, threats, opportunities, and constraints of the environment within the context of the organization's history." In this case, the strategy of IEVAP is to provide a banking system for the country of Iraq that allows the free flow of money with full personal and public accountability for all transactions within an environment still reeling from a history steeped in corruption.

56

### 3.     Transformation

At the center or heart of this transformation model is the organization itself.  The organization consists of work, people formal and informal organizations.  In this case, the organization is the country of Iraq.

- **Work**

Work "describe(s) the basic and inherent activity engaged in by the organization, its units, and its people in furthering the company's strategy" [28].  Because the organization in this model is a country, the "work" involves many different groups of people.  At the heart of the country are the government and its employees whose work consists of trying to rebuild Iraq from the ground up.  Having money used in the proper way and ensuring that money gets to the right people for the right reasons is paramount to the success of rebuilding the country of Iraq.  Every dollar that is stolen or misplaced is a dollar that could have prevented another improvised explosive device (IED) or been used to rebuild a school or hospital.  Fraud and financial corruption are serious roadblocks to the very important work that still needs to be done in Iraq in order for the country to thrive.

- **People**

The question is who are the people within this very unique organization?  Ultimately, they are the patriotic Iraqis who are willing to risk their lives today to have a better Iraq tomorrow.  Such patriots include government employees, as well as police and military personnel who patrol the streets.  The people who work these crucial jobs are already willing to risk their lives simply by their affiliation with the new Iraqi government.

- **Formal Organization**

The formal organization is defined as "the structures, systems, and processes that embody the patterns each organization creates to group people and the work they do and to coordinate their activity in ways designed to achieve the strategic objectives" [28].  In the country of Iraq, the formal organization ranges from the Prime Minister and his Cabinet to the leaders of the military.

- **Informal Organization**

The informal organization "encompasses a pattern of processes, practices, and political relationships that embodies the values, beliefs, and accepted behavioral norms of the individuals who work for the company" [28]. Because the cultural norms of Iraq are vastly different from those in the U.S., it is imperative to understand how those differences will affect the future implementation of the capability studied in this project. For example, in Iraq the people who have the most power are the sheiks, thus they are going to have the greatest influence on whether or not this new banking system is successful.

Also in Iraq, the colloquium that "cash is king" holds true. The only power these Sheiks have is the power to control their part of Iraq, which is currently done by force. Forceful control requires a certain number of people and hiring people requires money. How those people receive their money is an important factor. If this banking system leads to Sheiks being removed from the financial loop, it could create greater problems than the ones it is being designed to alleviate.

### 4. Output

In the end, "the ultimate purpose of the enterprise is to produce output—the pattern of activities, behavior, and performance of the system" [28]. In this model, the output consists of the system, the unit and the individual.

- **System**

The system refers to "*The total system.* The output measured in terms of goods and services produced, revenues, profits, shareholder return, job creation, community impact, and so on" [28]. In this case, the new banking system will allow the employees and the contractors to be paid with minimal amounts of money being lost due to fraud, thus pumping billions of dollars into the economy of Iraq. The more money there is in the economy the less likely people will turn to crime in order to make a living. Further, instead of the previous culture of corruption that pervaded Iraq, this banking system allows for a new culture of confidence and financial security.

- **Unit**

Units refer to "The performance and behavior of the various divisions, departments, and teams that make up the organization" [28]. This refers to the government as a whole.  In the case of the military, the military will not lose its soldiers for as much time because they will be able to conduct a number of financial transactions over the telephone.  Further, the more people who begin to bank, the easier it will be to make secure telephonic transactions.

- **Individual**

The individual refers to "the behavior, activities, and performance of the people within the organization" [28].  Positive results will occur in two areas: individuals will be less likely to steal because they know they are being tracked.  On the other hand, those officials who attempt to fraud the government will more easily be caught and removed from their positions.  Additionally, employees and contractors will be paid more quickly and with less inconvenience and threat to their personal time and safety**.**

## D.    FIT

Having discussed the congruence model as it relates to the Iraqi Banking Project and having set forth the desired results, the "fit" of this process must be examined in order to identify possible gaps in the solution.  Fit is defined as "the organization's performance [that] rests upon the alignment of each of the components—the work, people, structure, and operating environment—with all of the others" [28].  Finding the right "fit" is imperative to the success of this project, as every part of the Iraqi organization must learn to work together in order to achieve optimal results.

| Fit | The Issues |
|---|---|
| Individual—Organization | To what extent individual needs are met by the organizational arrangements. To what extent individuals hold clear or distorted perceptions of organizational structures, the convergence of individual and organizational goals. |
| Individual—Task | To what extent the needs of individuals are met by the tasks; to what extent individuals have skills and abilities to meet task demands. |
| Individual—Informal Organization | To what extent individual needs are met by the informal organization; to what extent does the informal organization make use of individual resources, consistent with informal goals. |
| Task—Organization | Whether the organizational arrangements are adequate to meet the demands of the task; whether organizational arrangements tend to motivate behavior consistent with task demands. |
| Task—Informal Organization | Whether the informal organization structure facilitates task performance or not; whether it hinders or promotes meeting the demands of the task. |
| Organization—Informal Organization | Whether the goals, rewards, and structures of the informal organization are consistent with those of the formal organization. |

Table 6.     Fit [From 28]

In Iraq, the group with the least amount of fit is the informal organization.  This fact will become more evident when the individuals that will be the greatest resisters to change are discussed.  Much of the problem in the informal organization with regards to "fit" is based on a lack of readiness for change.

**E.     ASSESSING A READINESS FOR CHANGE**

Now that the problem has been diagnosed and the fit has been accessed, the next step is to asses the Iraqi's readiness for change.  The change equation developed by Dr. Michael A. Beer will be used to make that assessment.  This equation is not a mathematical equation; it is simply a theoretical equation stated mathematically as:

Amount of Change = (Dissatisfaction X Model X Process) > Cost of Change

Simply put, the amount of change that is desired must be equal to the product of the dissatisfaction for the way things are; a model to bring about that change; and the process of implementing it. All of that must be greater than the cost of the change.

### 1. Amount of Change

The amount of change refers to how much change is actually desired. In this case, the change is large because it is asking the people of Iraq to modify the way they have done business for many, many years.

### 2. Dissatisfaction

Although there is no real way to assess the level of dissatisfaction the employees of Iraq are currently experiencing, because of the corruption and danger involved in being a part of the federal government, it can assume with some certainty that the Iraqi government employees are less than satisfied. Despite this fact, these employees are also extremely skeptical of change. In order to combat this resistance to change, the Prime Minister of Iraq will need to create some sort of buy-in for members of both the formal and informal organizations. "Buy-in" is the process of convincing the employees of the Iraqi government, through education, incentive programs and improved working conditions, that a new system of financial responsibility is worth the effort. The burden for convincing the people of Iraq that change is not only necessary, but also beneficial rests on the shoulders of the government itself.

### 3. The Model

As Professor Michael Beer writes, "A vision of the future state of the organization, the behaviors and attitudes as well as the structure and systems, is required for change to occur" [29]. Notice that he does not state that this is required for a successful change to occur. Beer simply states that in order for a change to occur at all, a vision must be present. Unless the leaders of Iraq can offer their people a clear vision of a more positive future, it will be impossible for them to implement change.

### 4.  The Process

The process of implementing the necessary modifications in Iraq will follow J.P. Kotter's "Process of Renewing and Transforming organizations" [30]. This eight step process developed by Kotter will serve as a roadmap for successful change in the country. In addition, Schein's multistage cycle of "Unfreeze-Change-Refreeze" has been overlaid on J.P. Kotter's process. These two processes combine and reinforce the necessary process depicted below.

## The Process of Renewing and Transforming the Iraqi Banking System

**FREEZE**

1. **Establishing a Greater Sense of Urgency**
   – Getting people to examine seriously the competitive realities
   – Identify crises, potential crises, or major opportunities
2. **Creating a Guiding Coalition**
   – Putting together a group with enough power to lead the change
   – Getting the group to work together like a team
3. **Establishing a Transformational Vision**
   – Creating a vision to help direct the change effort
   – Developing strategies for achieving that vision

**CHANGE**

4. **Communicating the Change Vision**
   – Using every vehicle possible to constantly communicate the new vision and strategies.
   – Role modeling needed behavior by the guiding coalition
5. **Empowering Others to Act**
   – Getting rid of blockers
   – Changing systems or structures that seriously undermine the change vision
   – Encouraging risk taking and nontraditional ideas, activities, and actions

**UNFREEZING**

6. **Creating Short-Term Wins**
   – Planning for some visible performance improvements
   – Creating those wins
   – Visibly recognizing and rewarding people who made the wins possible
7. **Consolidating Gains and Producing Even More Change**
   – Using increased credibility to change all systems, structures, and policies that don't fit together and don't fit the transformation vision
   – Hiring, promoting and developing people who can implement the change vision
   – Reinvigorating the process with new projects, themes and change agents
8. **Institutionalizing New Approaches into the Culture**
   – Creating better performance through customer and productivity oriented behavior, more and better leadership, and more effective management
   – Articulating the connections between behaviors and firm success
   – Developing means to ensure leadership development and succession
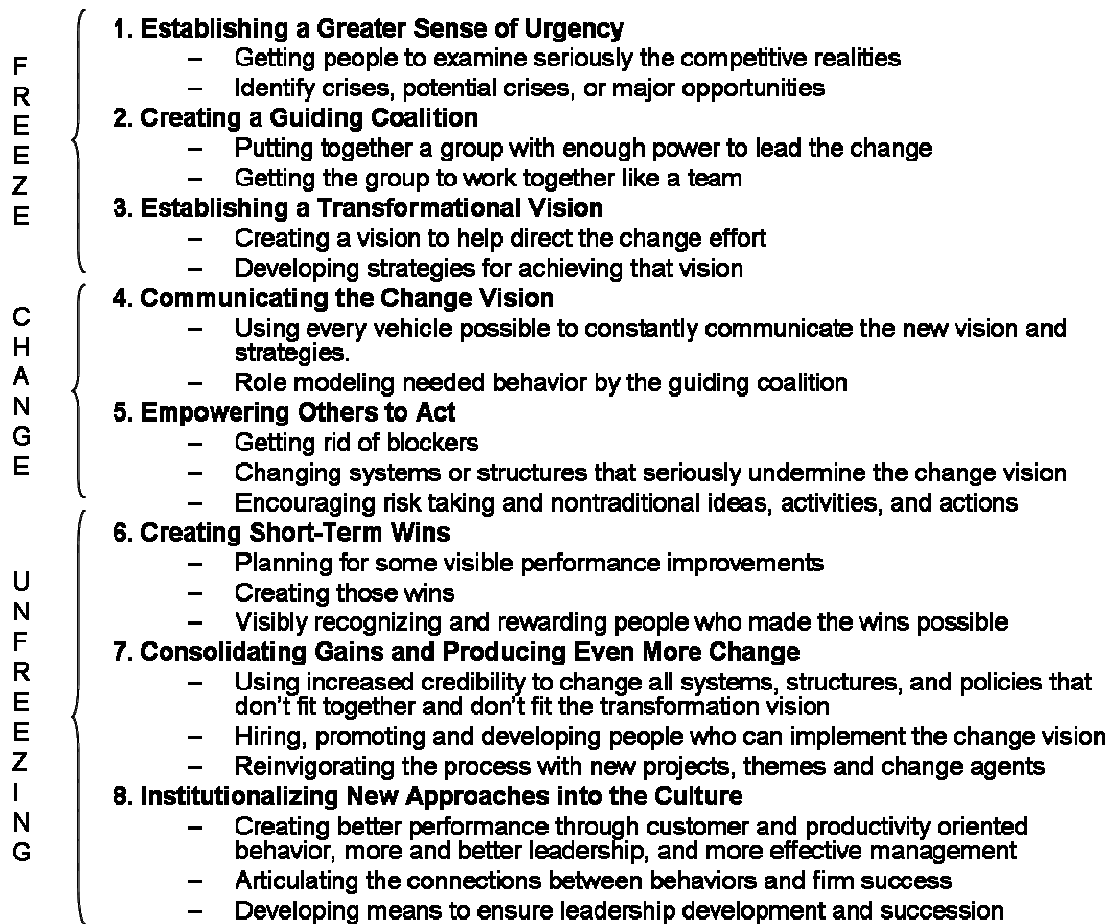
Figure 15.      The Process of Renewing and Transforming the Iraqi Banking System [After 30]

- **Establishing a Greater Sense of Urgency**

In a class taught at NPS, Professor Leonidas Doty states, "The most important aspect of change in an organization is a change in culture." The culture in Iraq is one that has suffered from years of oppression and corruption. According to William Bridges, "before you can begin something new, you have to end what used to be" [31]. If he is correct, then the corruption that has pervaded Iraqi society must be ended. Having the ability to hold people accountable is a key factor in ending corruption.

As mentioned previously, the American public is losing its patience with this war and there is talk of a pullout in 2008. At the very least, this change must be affected before the next president is inaugurated because of the inherent uncertainty that accompanies a shift in administration. Both the U.S. and Iraqi governments must work quickly if there is going to be significant progress. Chapter I lists the current schedule for the implementation of IEVAP. This schedule contains six phases. In order to expedite this process, the following schedule is recommended:

- **Phase 1.** Pilot menu-driven phone and laptop system and demonstration that voice authentication technology can work with sufficient accuracy.

  - **Phase 1A.** Develop and demonstrate a bilingual voice-activated menu-driven phone system in English and Arabic.

  - **Phase 1B.** Test and demonstrate speaker verification technology in English.

  - **Phase 1C.** Test and demonstrate speaker verification technology in Iraqi-Arabic.

- **Phase 2.** Detailed development of enrollment applications and preparation of systems/applications for deployment.

- **Phase 3.** Deployment and operational testing in Iraq.

- **Phase 4.** Broader deployment decision.

- **Creating the Guiding Coalition**

Because most of the problems in Iraq begin at the top and filter down, the top down method is recommended. The government of Iraq, beginning with the office of the Prime Minister, must acknowledge and embrace the required changes to include

prosecuting or relieving those in positions of authority that have misused funds intended for the rebuilding of their country.  The upper echelons of the Iraqi government must also accept and establish a new vision for a financially secure and responsible Iraq.

- **Establishing a Transformational Vision and Strategy**

As mentioned previously, the greatest amount of resistance is likely to come from the sheiks that are currently in power in a cash based society.  Schein states "Present behavior or attitudes must actually be disconfirmed, or must fail to be confirmed over a period of time" [32].  In Iraq, Nouri al-Maliki needs to show that he is concerned about the current levels of corruption infecting his country.  He must also make it very clear that the way things were under the previous regime is no longer acceptable and requires change.  He must offer the people of Iraq a new vision for a better Iraq.  According to Jick, this vision should "incorporate four elements: (1) customer orientation, (2) employee focus, (3) organizational competencies, and (4) standards of excellence" [32].

- **Communicating the Change Vision**

Kotter suggests that every vehicle possible for change be used.  In this case, the creators of the vision, i.e. the Iraqi and U.S. governments in coalition, have now become the "influencers" for that vision.  This step also correlates to Schein's Second Stage of "Changing."  Kotter and Schein both agree that this is the time in the change process to set up a definitive role model.  Schein states "One of the most powerful ways of learning a new point of view or concept or attitude is to see it in operation in another person and to use that person as a role model for one's own new attitude or behavior" [31].

Simply stated, Iraq needs someone they can look up to in the area of financial "freedom."  As they begin modeling the behavior of another organization that refuses to tolerate corruption and demands accountability for the use and distribution of funds, they will begin to learn and practice a new way of financial behavior.

- **Empowering Others to Act**

In this step, Kotter suggests that you must rid your organization of any blockers. In Iraq, the blockers are going to be those individuals who have the most to gain by keeping the old system in place, in other words those who are corrupt themselves. The Prime Minister needs to let the delayed investigations proceed and some ministers will likely need to be fired in order to set an example that this type of behavior is not tolerated. He needs to make it clear that with the new technology in place all financial transactions will be tracked. Anyone who makes an illegal transaction will be caught and subsequently prosecuted.

- **Creating Short-Term Wins**

Kotter states that there are opportunities for a visible performance improvement and rewarding those who make the wins possible. This is the beginning of Schein's refreezing process that allows for the solidification of change. As Iraq begins to solidify these changes, the government will begin to show the short-term gains of the new system in terms of money saved and illegal transactions prosecuted.

- **Consolidating Gains and Producing Even More Change**

At the onset of the refreezing process, those undergoing the change will "test" each other as Schein suggests. The "employees" will be leery about settling into their new environment, wondering if they can actually rely on this change to be a meaningful and lasting one. At the same time, the Iraqi government will want to see if these changes actually improve accountability and safety. Considering that the goal is to do better than the current loss of billions of dollars yearly, such a change should not be difficult to achieve. Schein states that this step "may require a good deal more give-and-take and thus may be initially slower but it will last longer" [31]. Because the goal is to have a long-term effective change in the financial situation in Iraq, this is the right approach to take.

- **Institutionalizing New Approaches into the Culture**

This step is the most tenuous time during a change. In Kotter's model, this is the step where the Iraqi Government will see if the change has made a positive impact in the area of financial accountability. If the change proves to be a success, it will open the country up to future change and growth with this system. If, however, the whole process does not produce success and the "employees" are not being paid in a way that meets their needs, there might be a huge outcry to switch back to the old system, which will cause the government to avoid technical changes in the future.

## 5. The Cost of Change

Initially, the greatest cost of change is going to be financial. As previously stated, ending financial corruption in Iraq is going to be a gradual process that will begin with the government, military and the agencies that do business with the government. The Iraqi government must realize that there exists a culture of corruption within their country and that the true cost of not changing could result in the loss of billions of dollars and a loss of credibility for the Iraqi government.

For the government employees, military personnel, and contractors this change will create a huge shift in the way they are paid. Because such a change affects their income, they may be resistant at first. At the end of the day however, if these employees see that they are still being paid in full and on time, but in a safer and more efficient manner, they will be duly satisfied. In order to achieve this level of satisfaction, however, the implementation of this strategy must be smooth and the benefits well publicized.

## F. A NOTE OF CAUTION

## 1. Archetypes

Archetypes are mental models that a person carries with them and are important to understand because "certain patterns of structure recur again and again" [27]. Although many archetypes apply to the problems in Iraq, the Archetype that will be discussed is "fixes that fail" [27].
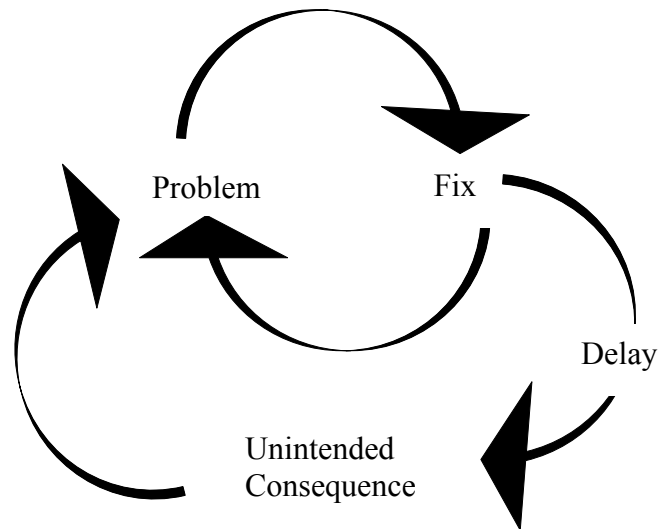
Figure 16.        Fixes that Fail [After 27]

### 2.        **Fixes that Fail**

In the fixes that fail archetype, a fix that is effective in the short term has consequences that are unforeseen and which ultimately require another fix.  This is much like the situation involving central payday locations for military personnel mentioned earlier in this chapter.  The Iraqi government tried to fix the problem of payroll corruption by consolidating the payment of its Armed Forces.  Unintentionally, this created other problems.  IEVAP contends to fix these and other problems with a telephonic banking system, but there will be resistance.  As previously stated, using the congruence model the element that presents itself as the least likely to "fit" with this solution is the informal organization within Iraq.

Fredrick Nietzsche once wrote that absolute power corrupts absolutely.  In Iraq, it seems that any power at all is something that many will fight and die to protect. Therefore, when implementing this system, it is important that the informal organizations of the sheiks, and to a lesser extent warlords and religious leaders, not be overlooked. Simply put, they will be the greatest resisters to this change.  That being said, close attention must be paid to these leaders and to their feelings regarding the new method of payment and ultimately to the banking system that will be born from it.

The Early Warning Symptom is that the fix works initially, then stops working. Senge states that in order to prevent this from happening, leaders have to focus on the long term solutions. As the diagram suggests, any delay in payments to contractors or ultimately sheiks not receiving the money they need to maintain order in their specific areas, could lead to the unintended consequence of people not utilizing the new system and demanding a return to Egypt, so to speak. This would be a costly and unfortunate mistake. As mentioned previously, getting buy in from this group of people from the onset will ultimately lead to the success of this system.

## G.     CONCLUSION

The country of Iraq currently has a problem with financial corruption and lack of accountability. These problems have resulted in the loss of billions of dollars and possibly the loss of lives. Money that could have been used to make the lives of the Iraqi people better has instead been misplaced or misappropriated. According to the change equation, Iraq is ready for a change. Although that change might not directly involve all of the Iraqi people, it will ultimately affect all Iraqis.

The people of Iraq have the necessary dissatisfaction, a goal from the United States Government, and a process ready for them to enact. The benefit of implementing this new system far outweighs the cost, though the cost is largely financial. If done successfully, however, this new system will actually produce greater financial gain in the long run. As noted previously, it will be vitally important to include the local sheiks as part of the introduction of this system. With a clear vision and a commitment to creating effective and lasting change, the country of Iraq, which is currently steeped in financial corruption, can not only improve payroll methods and hold government agencies financially accountable; it can ultimately be a country that is financially free.

# VII.  CONCLUSION

## A.  SUMMARY DISCUSSION

Speaker verification technology is a good biometric and a perfect fit for the current banking problem in Iraq.  Furthermore, the Iraqi-Arabic system created by Nuance is a viable form of biometric technology and a credible solution to countering the corruption and lack of accountability that exists within Iraq.  Voice biometrics uses existing infostructure (landline, cellular or VoIP), which means that as soon as the Nuance system is completed it could be attached to a banking system that would allow for implementation at a relatively low cost.  Another benefit of this technology is that it is less intrusive and invasive than fingerprinting or retinal scanning.  This is especially beneficial when dealing with sheiks or other high profile users that would prefer to not be man handled by the trainers or bank employees – something that is required for fingerprinting.  Most importantly, this system is relatively easy to use and will require little training time for the user, an important factor for a technical change of this magnitude.  In addition to the system provided by Nuance, the files retrieved by the system could be used by other systems for the purposes of voice identification.

This thesis documents the results of the independent test of Nuance by the NPS team's efforts in the conclusion of IEVAP Phase 1C.   In doing so, the NPS team successfully tested the claims made by Nuance concerning their speaker-verification system for Iraqi-Arabic.  The NPS test consisted of 41 native Iraqi speakers conducting enrollments with 1377 speaker verification attempts, 11 False Rejects and 1182 imposter trials, 59 False Accepts.  This resulted in a False Rejection Rate of .8% and False Acceptance Rate of 4.9%.  This yielded an accuracy of 97.3%.  The intent of this project was to validate the system for its future employment in the country of Iraq in order to revitalize the Iraqi banking system.

## B. RECOMMENDATIONS FOR FURTHER RESEARCH

At the end of IEVAP Phase 1C, it is clear that the research objectives have been met and this product should be developed to act as an entry point into a new banking system that will allow for increased integrity and accountability. It is recommended that Phase 2 of this project include the completion of this entry point and a proof of concept banking application if no commercial banking applications are available. This technology, however, is by no means limited to banking. Further research should be considered in the areas of:

– Conducting and/or building of a system that would allow for remote access for use at vehicle checkpoints and base entry points either independently or using reach back through IEEE802.11 or IEEE802.16.

– Building into the current Nuance System the capability to use Multi Factor Authentication to include context free voice recognition.

– Conducting tests to verify whether the voice recordings alone could be used for purposes of voice identification.

– Creating a proof of concept system for VIP entry into the green zone.

– Conducting tests to see if this technology could be used within the United States in support of the Department of Homeland Defense.

## C. FINAL THOUGHTS

The United States is currently enmeshed in a war with a nation abounding in complexities. Terrorism, though our most immediate concern, is not the only problem threatening the stability of Iraq. Financial corruption is also a huge concern and one that is costing both Iraq and America money and lives. After five years of war and more than thirty-five hundred U.S. lives, the people of America grow weary of our involvement. The time for helping Iraq attain independence is now, but the question of financial corruption must be addressed in order to achieve this goal. The current banking situation in Iraq is unacceptable and in dire need of effective change.

Nuance has created a front door to a banking system that will revolutionize the way business is conducted within the Iraqi government. No longer will billions of

dollars, both U.S. and Iraqi, be lost, stolen, misappropriated or squandered with no form of accountability.  Instead of money being used to line the pockets of the corrupt, it will be used as it was intended, for restoring the infrastructure of Iraq and giving the Iraqi people a safer, more stable, more financially free country.  As the people of Iraq begin to see these changes take hold, it is likely they will be less inclined to support anti-Iraqi forces and more inclined to work with their own government for the betterment of their nation.  Though seemingly an expensive investment, implementing the Nuance system will offer a return that is well worth the cost.  Not only will it save money and deter corruption, it will also save lives, both Iraqi and American and will take our country one-step closer to ending the Global War on Terrorism.

THIS PAGE INTENTIONALLY LEFT BLANK

# APPENDIX A.



**NUANCE PROFESSIONAL SERVICES** | Version 1.1  05/17/07

**Naval Postgraduate School**
**Iraqi Arabic Nuance Caller Authentication**

Performance Report

# Document History

| Author | Name |
|---|---|
| Speech Scientist | Ramon Prieto |
| Solutions Manager | Krishnan Srinivasan / Charles Jankowski |
| Project Manager | Ramon Prieto |

| Application | | | |
|---|---|---|---|
| Description | | This report contains the details of the Performance of the Iraqi Arabic Nuance Caller Authentication Pilot developed for the Naval Postgraduate School. The application allows a user to enroll his voiceprint by saying digits 1-9. It also verifies the user's voice or rejects it depending on the user being an impostor or a claimant. | |

| Date | Revised by | Version | Summary of Changes |
|---|---|---|---|
| 05/17/2007 | | 1.1 | Initial Version |

74

# TABLE OF CONTENTS

75

# Chapter 1: Overview

This report analyzes the performance of the Iraqi Arabic Nuance Caller Authentication system, on a Pilot executed in Jordan with native Iraqi Arabic Speakers from several dialects.

The application consists on an Iraqi Arabic Localization of the Nuance Caller Authentication Product. Prompts, Acoustic Models, Speaker Verification Models, Grammars, and the Nuance Caller Authentication engine were localized accordingly for the realization of this project. The localization took place between November 2007 and May 2007.

The pilot calls were executed between March 27th, 2007 and May 16th, 2007. A total of 2355 life verification calls took place on this pilot during those dates, on a total of 239 subjects, which enrolled 1 voiceprint each. A total of 11775 impostor calls were simulated by running several verification calls already recorded on voiceprints that did not correspond to the caller. Impostor calls were simulated without using a life environment. 239 voiceprints and subjects were also used for the impostor trials. The total number of calls is of: 14320.

## 1.1    Executive Summary

### 1.1.1  Performance

The performance of the Iraqi Arabic Nuance Caller Authentication application was analyzed from 3 different perspectives. These perspectives correspond to:
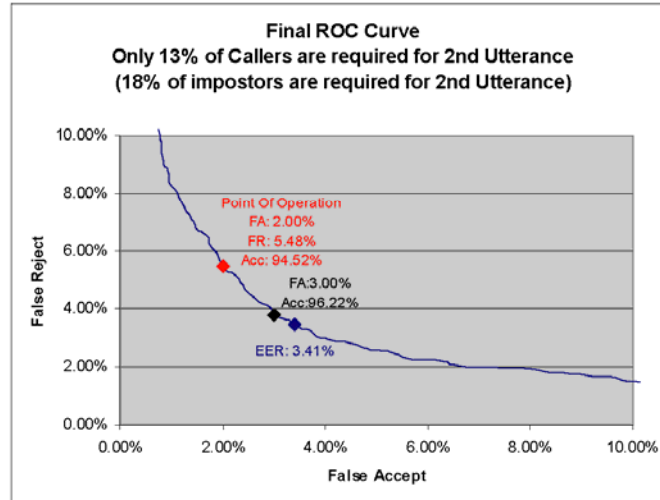
Equal Error Rate

- The equal Error rate achieved in the final recommended application is of: **3.41% EER**. This is a very encouraging Equal Error Rate for a real life application and definitely higher than expected for the development of a new language on a speaker verification application. The equal error rate reported by Nuance Communications on an American English NCA application for the Naval Postgraduate School was 3.0% [1]. The ROC curve and EER of the Iraqi Arabic NCA application can be seen in the figure below.

Speaker Verification Accuracy

- Speaker Verification accuracy, also known as the accuracy of the system, was evaluated on 2 different points in the ROC curve. The system got a total accuracy of **94.52% at a 2.00%** false acceptance rate, and it got a **96.22% accuracy at a 3.00%** false acceptance rate. The ROC curve and EER of the application can be seen in the figure below.

Speech Recognition Accuracy

- To be able to deploy an Iraqi Arabic application, the development of Iraqi Arabic acoustic models (also known as Speech Recognition Models) had to be executed. There were significant improvements achieved in speech recognition accuracy in comparison to the original Jordanian Arabic models. The recognition of length 7 digit strings improved to **94.87%** from 71.28%.The yes/no accuracy improved to **90.31%** from 64.29%. Our new Iraqi Arabic models performed recognition of length 4 PIN digits with an improvement to **80.95%** from an original 45.22%.

76

**Final ROC Curve**
**Only 13% of Callers are required for 2nd Utterance**
**(18% of impostors are required for 2nd Utterance)**



### 1.1.2  Application Characteristics

The final application delivered to NPS consists of a dialog state that can operate in 2 tasks. The first task, or enrollment task is performed by requesting the user to say his account number, request the user to confirm if he wants to be enrolled and finally request the user to pronounce digits 1-9, 3 different times.

The second task, or verification task consists in requesting the user, given that he has already enrolled, his account number. Then, the user is requested to pronounce digits 1-9. For the sake of collecting as much data as possible, the application requested almost all users to pronounce digits 1-9 two different times. However, the final recommended system allow for achieving the EER, Accuracy and ROC curves of the figure above by **requesting for a second utterance 1-9 ONLY to 13.08%** of the callers. See figure below for details.

77
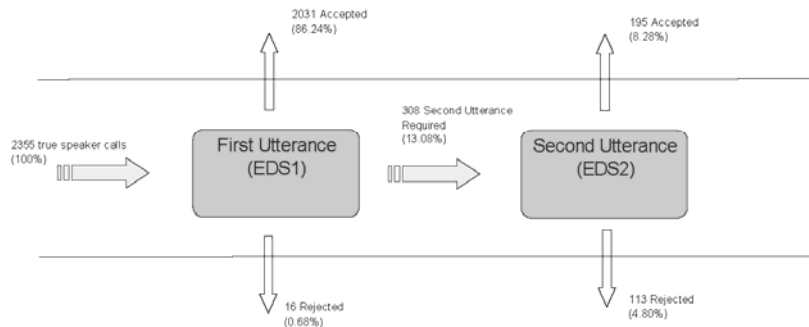
# Chapter 2: Data Collections and Pilot

*This section describes the data collections that had to be performed to both develop the application and evaluate the final accuracy of the system.*

## 2.1 Training Data Collection

Before the execution of this project, Iraqi Arabic Acoustic Models (Speech Recognition Models) and Iraqi Arabic Speaker Identification models were not available for deployment. To be able to execute the project and evaluate the performance of the Iraqi Arabic NCA system, new Iraqi Arabic Acoustic Models and new Iraqi Arabic Speaker Verification models had to be trained.

To be able to train them, data collections of 200 different speakers, that natively speak Iraqi Arabic had to be collected. Nuance selected a partner to execute a training data collection of 204 different Native Iraqi Arabic speakers.

Each speaker was requested to speak 120 different utterances in two sessions. The first session would correspond to a cellphone handheld session, and the second session would correspond to a landline handheld session. 60 utterances were pronounced by the speaker on each session. To be able to cover all accents, genders and channels Nuance accounted for the following distributions: 65% Male speakers and 36% female speakers. See figure below.



Gender Distribution: 204 People in Total

36%  64%

- Male
- Female

78
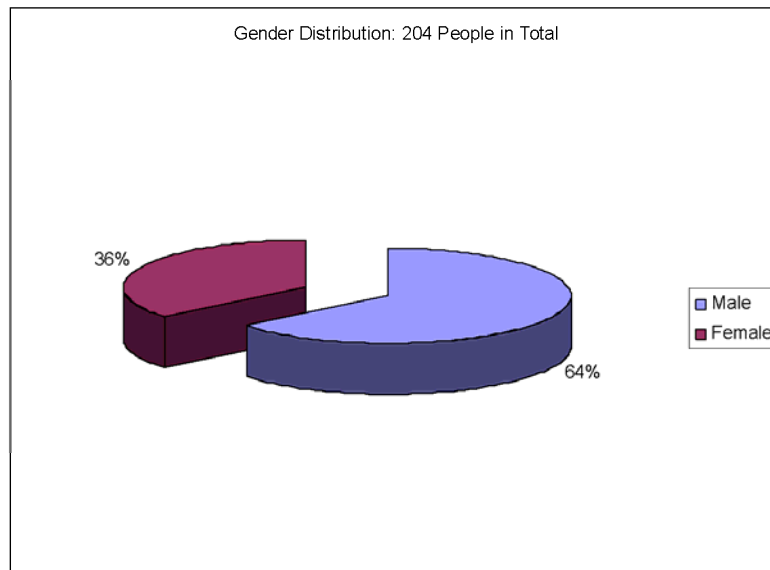
In each call, each user would say 60 different utterances. The user was requested to say 15 different short commands, 15 length 7 digit sequences and 30 1-9 sequences. To account for dialect variability on the way people says short commands and digits, the population of 204 training subjects was broken down across Baghdadi Iraqi Arabic Dialect, Northern Iraqi Arabic dialect and Other dialects. The percentage of each dialect on the population is described in the figure below.



Dialect Distribution: 204 People in Total
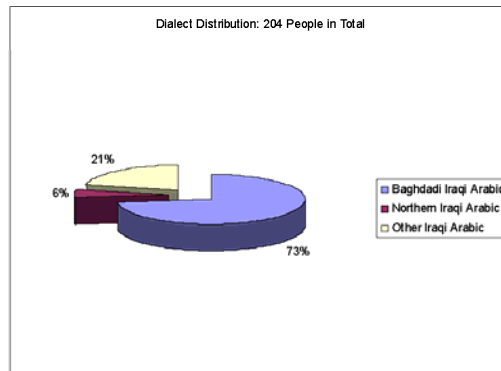
As the user was requested to make two calls, one from a cell phone and one from a landline phone, the distribution between different channels was very close to 50%/50%. The details of the landline/cell phone distribution can be seen below.



Channel Distribution: 407 Calls in Total

79
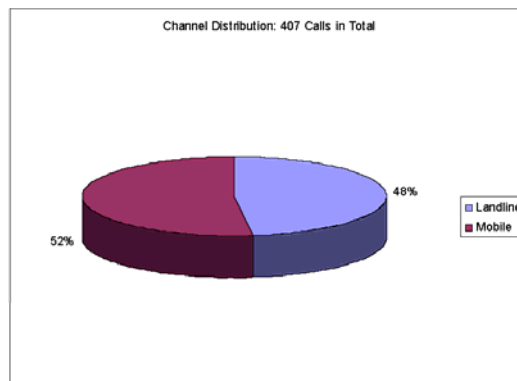
## 2.2    Pilot Data Collection

To be able to measure the performance of the application, there was the need of having real callers calling the application in a real life setting. The application was hosted at Nuance Communications on the number (650) 289 8447. This number allowed for 6 concurrent callers calling the system, on 6 concurrent calls, on the same phone number. Subjects were scheduled to call the system from Jordan originally.  239 subjects living in Jordan that spoke natively the Iraqi Arabic language were recruited to make the experiments.

The data collection took place in two different phases. In phase I, callers would enroll and make 5 different verification calls. In phase II, callers would make 5 additional verification calls.  The time gap between Phase II and Phase I was originally scheduled to be 2 weeks. However, complications in recruiting people back in Jordan for Phase II delayed the starting time of Phase II and the time gap between Phase I and Phase II was 1 month and 2 weeks (first callers came on March 27th for Phase I. First callers came on May 9th for Phase II.

To account for dialect distribution in the data collection and have enough subjects from all dialectal regions from Iraq, the following dialect distribution was planned among subjects of this pilot data collection:
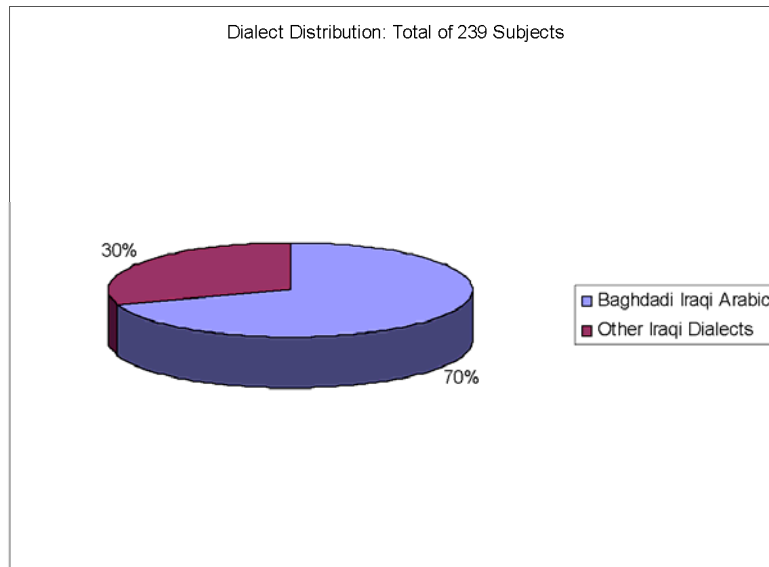


From our estimates, out of the 239 subjects, around 145 completed Phase I and Phase II calling from Jordan, with 1 month and 2 weeks of time gap between phases. Around 10 Callers completed Phase I and Phase II making calls from Australia, with a 3 days gap. Around 45 subjects completed Phase I and Phase II in Jordan with a 3 days gap. And around 39 speakers attended only Phase I in Jordan without completing Phase II.

80

The need of a gender distribution as close to 50%/50% is necessary. However, culturally, in countries in the middle east It is significantly harder to recruit female subjects than male subjects. As a result the intended gender distribution was planned to be the following:

Gender Distribution: Total of 239 Subjects

30%

70%

☐ Male
■ Female
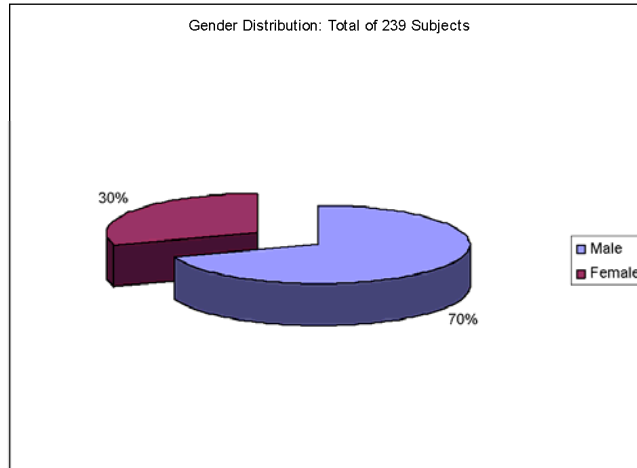
In terms of channel calls, the data collection was planned to have 85% of speakers enrolling from cellphone. 15% of speakers enrolling from landline. 80% of verification calls made from the same channel the user enrolled on, and 20% of verification calls made from a different channel the user enrolled on. As a result, the total channel distrubtion of calls, was plan to be the following:

Channel Distribution: Total of 14130 calls

27%

73%

☐ Cellphone
■ Landline

81
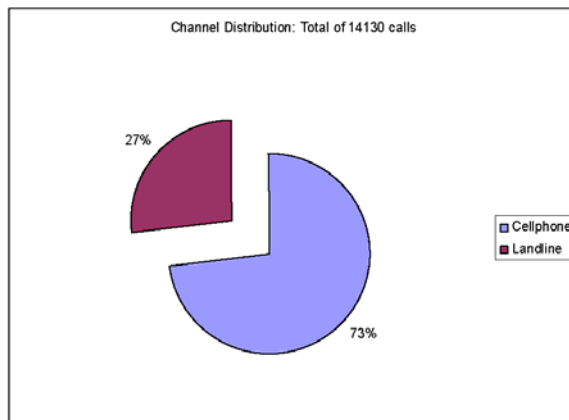
The chart below shows the number of calls made into the pilot system from the start of the pilot in March 27 2007 to the end of the pilot that finally took place on May 16th 2007. We can see that the major flow of calls occurred between March 27th , April the 7th, and May 9th to May 16th.  The very few calls made in between April the 25th  and May the 3rd were calls made from Australia instead of Jordan
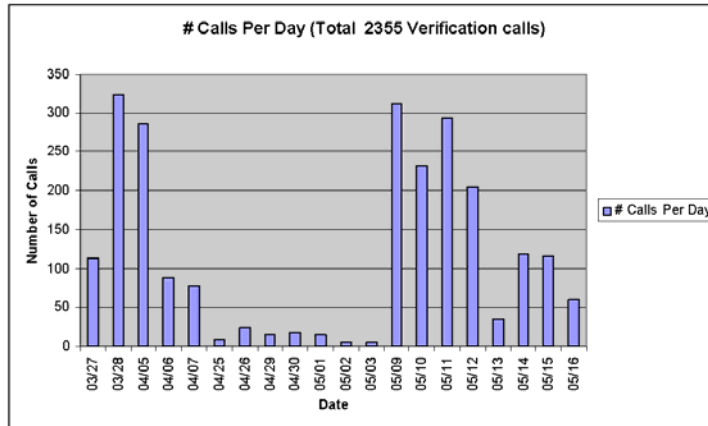


The chart below shows the number of total calls, including verification and simulated impostor calls from which we derived the results on section:  "Application Performance". Both the chart above and the chart below do not include enrolment calls.

82

83
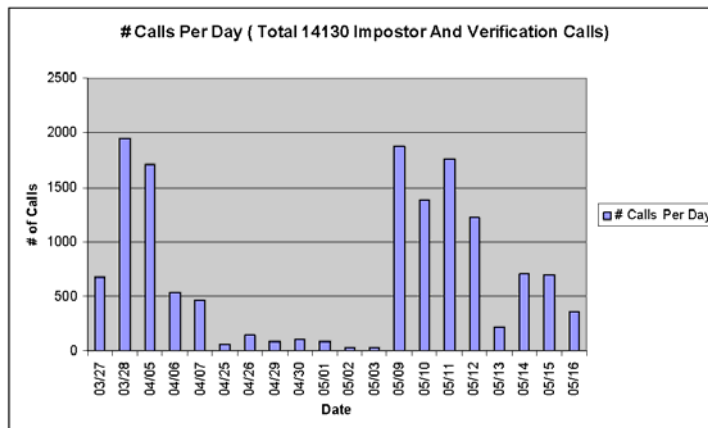
The number of enrolled users per date is shown on the chart below. Although the total number of subjects enrolled on the system is 263, only 239 were used in the experiments. The users that make the difference were detected to be unintended impostors between phase I and phase II and users whose waveform had tonal noise in their enrolment or verification utterances. We can see that most of the enrolment calls took place in March and April. We can also see that enrolments took place later in the data collection only for replacements of the subjects that, although they attended Phase I, they did not attend Phase II. For these subjects, which were recorded mostly in Jordan and some in Australia, there was a 3 day gap between Phase I and Phase II.



Enrollment Calls Per Date
Total 263 subjects. Only 239 Used on Evaluation.

83
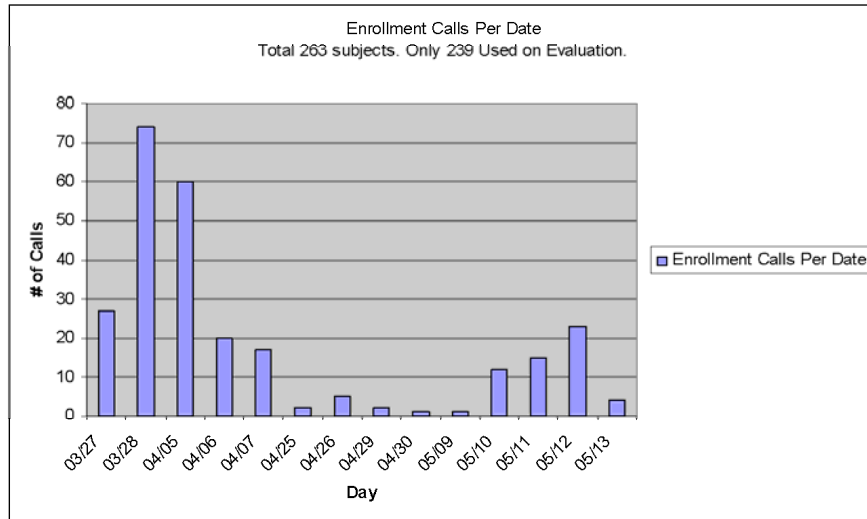
# Chapter 3:  Application Development

Several steps had to take place in order to develop the Iraqi Arabic Nuance Caller Authentication System. This section describes the different parts that had to be developed.
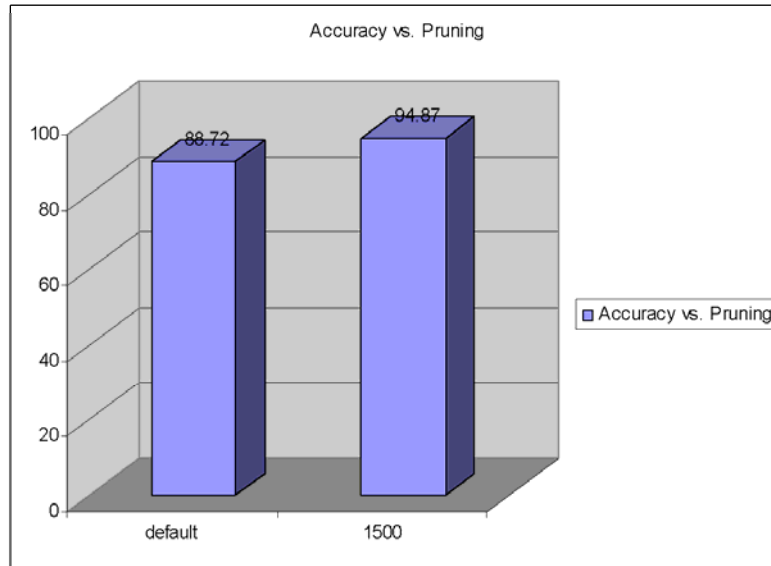
## 3.1    Iraqi Arabic Adaptation of Acoustic Models

We used the training data collection described in section: "Training Data Collection" in order to develop the Speech Recognition Acoustic Models (or Speech Recognition Models) for Iraqi Arabic. The main objective was to improve Speech Recognition performance over the Jordanian Arabic speech recognition models.

We used the training data to adapt the current Jordanian Models. To adapt them we used several smoothing coefficients. These are 500, 2000, and 5000. We also used 2 different dictionaries: a dictionary with only a minimal set of pronunciations for digits, and a dictionary with all possible pronunciations for digits. The table below shows the speech recognition performance, at 0 rejection rate, of a length 7 digit sequence. We can see that the system adapted with coefficient 500, using the large dictionary performs the best, at a 94.87% accuracy. This is an important improvement compared to the Jordanian models, which could only get a 71,28% accuracy. The test with the results below were not done on life data. We used a subset of the training speakers to develop a test set. None of the speakers in the test set are in the training set.

84
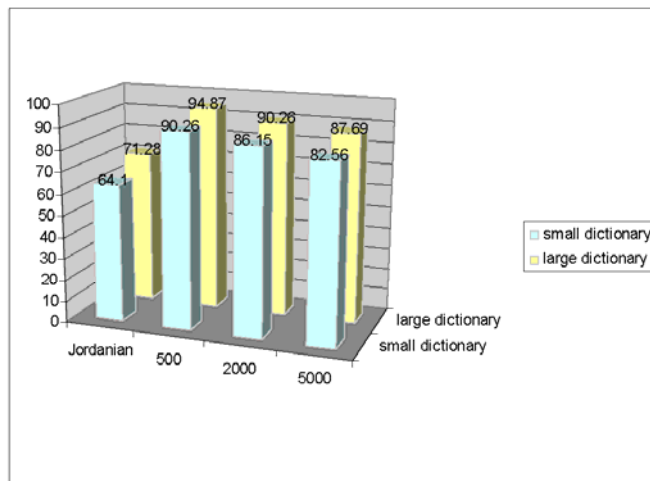
In addition to experiments using different dictionaries and adaptation coefficients, we also modified several of the speech recognition parameters to analyze improvements. From all the parameters, the parameter with the biggest contribution on speech recognition performance was the Pruning parameter. We can see that increasing the pruning value from the default value to 1500 brought significant improvements, from 88.7% to 94.87%. We can see that in the chart below.



## 3.2    Development of Speaker Verification Models

We used the training data collection described in section: "Training Data Collection" in order to develop the Speaker Verification Models. The main objective was to develop several Speaker Verification models and select the one with the best performance to deliver it to NPS. Since we had to develop these models before the full pilot data collection was finished, we made several versions of Speaker Verification calls using the training data collection, we selected a small set of calls from the pilot data collection completed so far and compared their performance on these small set of calls based on EER.

We compared their performance based on using 2 utterances per verification/impostor trial.

85
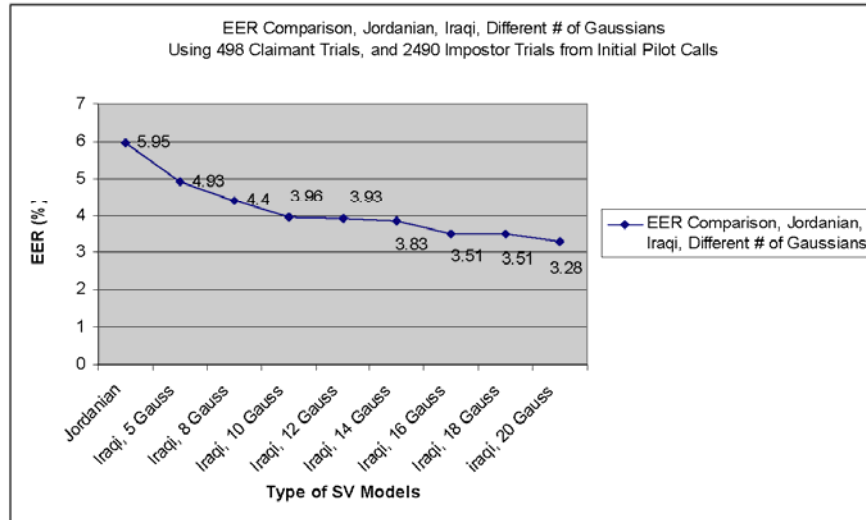
We can see in the chart above that the performance considerably improved as we used SV models trained on Iraqi data instead of SV models trained on Jordanian Data. Furthermore, using more Gaussian models in the SV models improved the performance from 4.93% (using 5 Gaussians per phoneme) to 3.28% EER (Using 20 Gaussians per phoneme).

The SV models delivered to NPS for their internal tests are based on Iraqi Training data, 20 Gaussians. Notice that the test set for these results is different than the one used for measuring the final performance of the system. In the chart above, since the experiments were done weeks before the end of the pilot data collection, we used only 498 Claimant trials and 2490 impostor trials from the first callers that called the pilot application in Jordan.

## 3.3     Prompt translation and Recording

As part of the NCA localization effort, the English prompts corresponding to the English NCA application had to be translated to Iraqi Arabic and then recorded properly. For this, our partner chose a semi professional voice and recorded the prompts in a recording studio. The prompts were delivered to NPS as part of the Iraqi Arabic NCA installation.

86

## 3.4 Grammar Translation

As part of the NCA localization effort, the English grammars corresponding to the English NCA application had to be translated to Iraqi Arabic and tested properly. For this, our partner used native Iraqi Arabic speakers as well as experience writing Nuance GSL grammars to properly translate the English grammars into Iraqi Arabic ones. These grammars were delivered to NPS as part of the Iraqi Arabic NCA installation.

## 3.5 Nuance Caller Authentication Localization

For the system to be properly installed, it had to be integrated and tested. The new Iraqi Arabic acoustic (Speech Recognition) models were integrated into the NCA engine, together with the Iraqi Arabic Speaker Verification models, the Iraqi Arabic Nuance Caller Authentication Prompts and the Iraqi Arabic Nuance Caller Authentication Grammars. All these deliverables integrated together had to be tested. We had our partner to hire several native Iraqi Arabic speakers to test the system under the several scenarios that the application allows for. Bugs were corrected, and the system was delivered to NPS.

# Chapter 4:   Application Performance

The numbers on this section of this report apply to the performance evaluation of the Iraqi Arabic Nuance Caller Authentication system on the Pilot data collection developed in Jordan, with native Iraqi Arabic speakers.

## 4.1    EER and Accuracy

According to the Nuance Voice Platform documentation, the definitions of EER, FAR, FRR and Accuracy are the following:
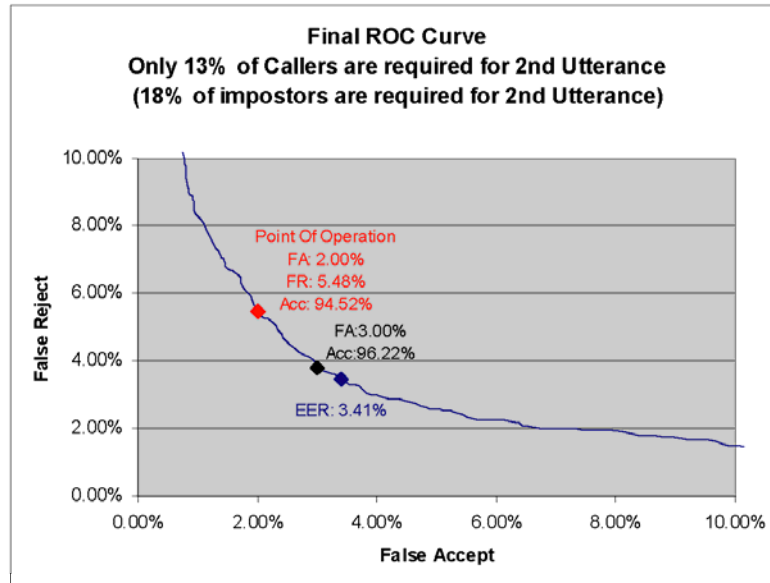
"

- **False acceptance (FA) rate**—The probability that an imposter is accepted into the application. **Note that the FA rate *is not* the percentage of calls** that result in a false acceptance, since this assumes that a large majority of callers are true speakers. The FA rate is the chance of being accepted given that you are an imposter. For example, a 1.0% FA rate does not mean that 1.0% of the total calls will be falsely accepted; it means that 1.0% of the imposters will be falsely accepted by the application. The total percentage of calls that result in a false acceptance is therefore equal to the FA rate multiplied by the probability that a caller is an imposter.
- **False rejection (FR) rate**—The probability that a true speaker is rejected by the application. It is assumed that almost all callers are true speakers; therefore, the FR rate **should be close** to the percentage of all calls that result in a false rejection.
- **Reprompt rate**—The probability that a caller is prompted for an additional utterances, when variable-length verification is turned on.

Verification accuracy is measured along a curve, called the *receiver operation curve* (ROC), that maps the FA rate and the FR rate pairs that can be achievable for an application (see diagram below). It is critical to understand that verification performance can only be specified by noting the FA rate and the corresponding FR rate at the same threshold.

The application can operate anywhere on the ROC curve. The location of the operating point on the curve is dictated by the verification thresholds required for your application. You modify the verification performance thresholds for your application by choosing a different operating point (a different FA rate/FR rate combination). As you decrease the FA rate, it is more difficult to get into the application, but the FR rate increases.

"

The resulting ROC that describes the performance of our application is displayed in the figure below. We can see that accuracies at tolerable FA rates are all above 90%. We can also see that accuracies of FA rates below 3% and below 2% are around 95% or higher.

88

**Final ROC Curve**
**Only 13% of Callers are required for 2nd Utterance**
**(18% of impostors are required for 2nd Utterance)**

Point Of Operation
FA: 2.00%
FR: 5.48%
Acc: 94.52%

FA:3.00%
Acc:96.22%

EER: 3.41%

The resulting EER on the operation curve above, is a very encouraging 3.41%. Well in the ballpark of EERs of languages that Nuance Provides. While more training data would be able to get us better numbers, these numbers are definitely desirable and recommended for use in industry.

Since in our Pilot data collection we allowed most of our callers to pronounce two utterances 1-9, we had the opportunity to measure what is the difference between using 1 utterance and 2 utterances on a single verification trial. The figure below shows the ROC curve for using all available utterances per trial. We can see that our EER and accuracies using all uterances available is of 3.41%, the same as using 2 utterances only on 13% of the population, and 1 utterance on the rest. We can also see that accuracies and FA rates are very close and similar.

89

ROC Curve
Using 2 Utterances Per Trial
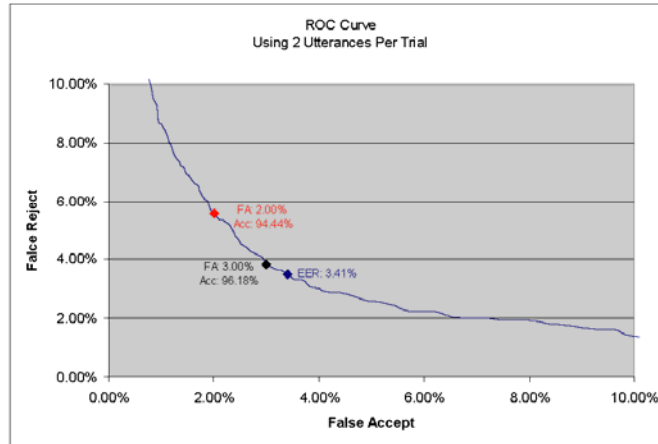
Looking at the ROC curve below, using only 1 utterance per verification trial and comparing it to the ROC curve of using 2 utterances per verification trial, we can see that, although there are differences in performance, using only 1 utterance per verification trial gives still an excellent performance ,with an EER below 4% and accuracies at around 95% or above.



ROC Curve
Using Only 1 Utterance Per Trial

90
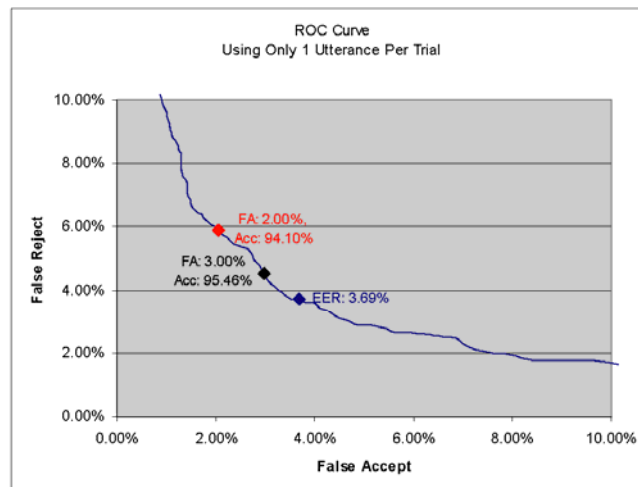
Finally, comparing accuracies of both phases of the pilot data collection, we can see that Phase I trials have a significantly better Accuracy and Equal Error Rate.
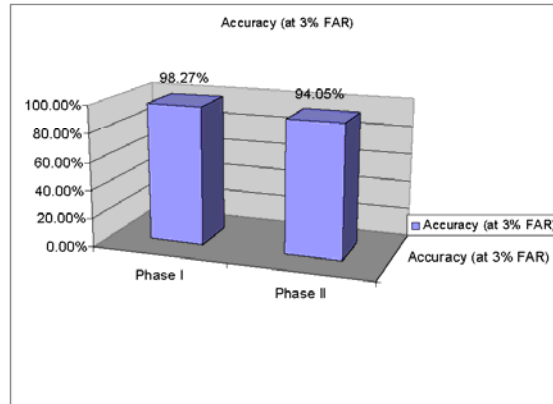


While the accuracy and EER of Phase I is significantly better than phase II, at 3%EER, the Iraqi Arabic Nuance Caller Authentication system is still able to perform, in Phase II, at accuracies around 95% and an excellent EER of around 4%.

91
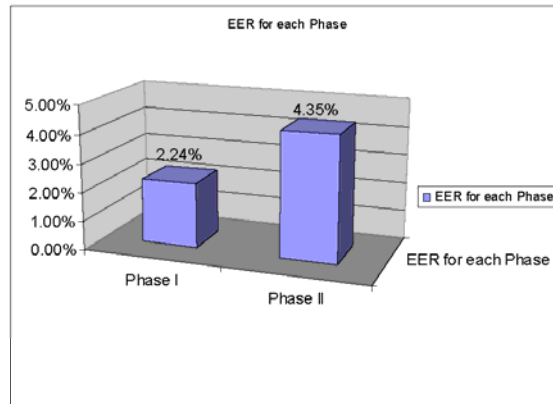
## 4.2 Number of Utterances per Trial

While more utterances used per verification trial give a better EER and better accuracies for the system, it's imperative that we keep the user friendliness of the system as high as possible without hurting performance. As a result we need to set the confidence thresholds of the NCA application in such a way that we are able to keep an EER and accuracy as close as if we were using 2 utterances per trial, but without asking the user to say 2 utterance unless it's strictly necessary.
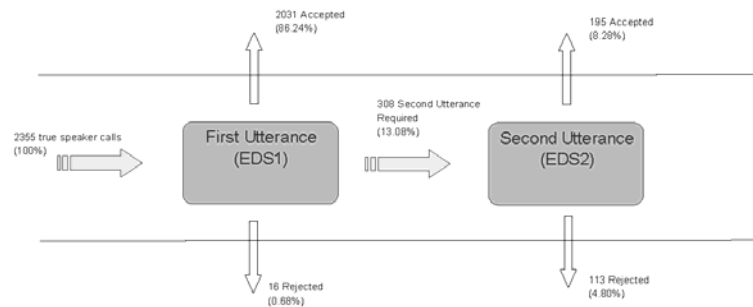


We found an excellent trade of for using only 1 utterance most of the time, requiring only 13.08% of the callers to pronounce a second utterance, and keeping the accuracy and EER as good as if we were using 2 utterances per trial. The flow and percentage of people that are required 2 utterances, accepted and rejected at each utterance can be seen in the figure above.

In the figure below we can see what happens to an impostor population. Since the impostor population is a tiny compared to the claimant population, we can allow a higher percentage to be requested a second utterance. However we are able to get outstanding EERs and accuracies by requiring only 18.21% of the impostor population a 2nd utterance.

92
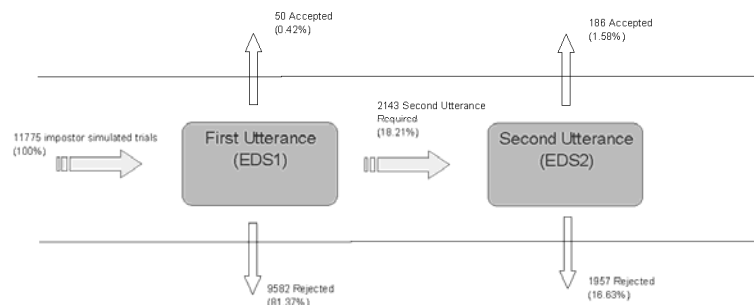
The confidence intervals used on NCA to achieve the results above (both ROC curves and percentage of people required a second utterance) are:

```
<eds1-verification-accept-threshold>54</eds1-verification-accept-threshold>
  - <!--
  Reject threshold for first utterance of verification
  -->
<eds1-verification-reject-threshold>43</eds1-verification-reject-threshold>
  - <!--
  Accept threshold for second utterance of verification
  -->
<eds2-verification-accept-threshold>50</eds2-verification-accept-threshold>
  - <!--
  Reject threshold for second utterance of verification (2% FA Rate)
  -->
<eds2-verification-reject-threshold>50</eds2-verification-reject-threshold>
```

93

# Chapter 5: Limitations of the System

It is important to mention that the system is not intended to work on universal situations. The situations enumerated below are situations in which the system has not been tested. The results of tests in those scenarios might be outstanding, good, medium or bad. Only the appropriate tests would determine the applicability of the Iraqi Arabic Nuance Caller Authentication system to those specific scenarios, which are:

- The system has not been tested with Arabic speakers of any other region than Iraq.

- The system has not been tested with Speakerphone calls.

- The system has not been tested with speakers of other languages but Arabic,.

- The system has not been tested with Bluetooth headsets.

- The system has not been tested with any other headsets or handsfree kits.

- The system has not been tested with Iraqi Arabic Speakers that have been living enough years outside Iraq as to loose the dialect or accent that is specific to Iraqi Arabic.

- The system has not been tested with cellular networks other than Jordanian Networks (which could be assumed to be close to an Iraqi Cellular Network).

- The System has not been tested with landline networks other than the Jordanian landline networks (which could be assumed to be close to an Iraqi Landline Network).

- The system has not been tested with other transmission channels, such as Voice Over IP, Iridium or satellite networks.

- The system has not been tested in heavy noisy environments, such as a battlefield, heavy car noise, voice or music in the background.

94

# Chapter 6:    Conclusions

1. Excellent performance, with 3.41% EER .

2. Phase II performance is not as good as Phase I, but can be easily addressed by using voiceprint adaptation. Voice print adaptation was not used on our pilot. (Typically, adaptation brings 25% relative FR reduction after 3 calls,  35% EER relative FR reduction after 6 calls). ([1]).

3. Pilot data collection is statistically significant, using more than 200 speakers, more than 2000 claimant trials, and more than 10,000 impostor trials. Speakers completed 2 different phases, and noisy utterances and unintended impostors were taken out of the samples.

4. Same methodology can be used to develop and test a Nuance Caller Authentication system that is able to do Speaker Verification for other languages, such as Dari, Pashto or Farsi.

95

# Chapter 7: References

[1] Charles Jankowski, "Naval Postgraduate School Verification Test Report"

# APPENDIX B

Major Jeffrey W. Withee and
Capt Edwin D. Pena
**Information Technology Management Curriculum**
**Naval Postgraduate School**
**Monterey, California 93943**

831-917-0073
edpena@nps.edu

To:     Protection of Human Subjects Committee

Subject:     Application for Human Subjects Review (Title): Testing and evaluation of Iraqi Enrollment Via Voice Authentication Project (IEVAP) in support of banking applications in Iraq.

1. Attached is a set of documents outlining a proposed experiment to be conducted over the next 3 to 4 months for our thesis research.

2. We are requesting approval of the described experimental protocol. An experimental outline is included for your reference that describes the methods and measures we plan to use.

3. We include the consent forms, privacy act statements, all materials and forms that a subject will read or fill-out, and the debriefing forms (if applicable) we will be using in the experiment.  Additionally, these forms will be provided in Arabic to any participant that requests it or an interpreter will be provided to translate the document for them.

4. We understand that any modifications to the protocol or instruments/measures will require submission of updated IRB paperwork and possible re-review.  Similarly, we understand that any untoward event or injury that involves a research participant will be reported immediately to the IRB Chair and NPS Dean of Research.

**J.W. Withee and E.D. Pena**

| APPLICATION FOR HUMAN SUBJECTS REVIEW (HSR) | HSR NUMBER *(to be assigned)* |
|---|---|

**PRINCIPAL INVESTIGATOR(S)** *(Full Name, Code, Telephone)*
Maj. JW Withee, USMC and Capt. E.D. Pena, USMC

**APPROVAL REQUESTED**      [X] New      [ ] Renewal

**LEVEL OF RISK**    [ ] Exempt    [X] Minimal    [ ] More than Minimal
Justification:  Dialing a telephone number and answering a few voice prompts poses little physical risk.  Additionally, no identifying information will be published as part of this study that may expose the participants to any risk.

| WORK WILL BE DONE IN (Site/Bldg/Rm) NPS Campus and one other site TBD outside of campus. | ESTIMATED NUMBER OF DAYS TO COMPLETE    45 Days. |
|---|---|
| MAXIMUM NUMBER OF SUBJECTS 100 | ESTIMATED LENGTH OF EACH SUBJECT'S PARTICIPATION:  Most will participate for, at most, an hour.  10 will participate up to 3 hours. |

**SPECIAL POPULATIONS THAT WILL BE USED AS SUBJECTS**
[X ] Subordinates    [ ] Minors    [ ] NPS Students    [ ] Special Needs (e.g. Pregnant women)

Specify safeguards to avoid undue influence and protect subject's rights:
Two computers will be used and personal information will be erased following the study.  A separate sheet will track participants; each person will be assigned an individual number.

**OUTSIDE COOPERATING INVESTIGATORS AND AGENCIES**
N/A

[  ] A copy of the cooperating institution's HSR decision is attached.

**TITLE OF EXPERIMENT AND DESCRIPTION OF RESEARCH** (attach additional sheet if needed).  Please see attached sheet.

I have read and understand NPS Notice on the Protection of Human Subjects. If there are any changes in any of the above information or any changes to the attached Protocol, Consent Form, or Debriefing Statement, I will suspend the experiment until I obtain new Committee approval.

SIGNATURE_____  DATE_____
SIGNATURE_____  DATE_____

**MINIMAL RISK CONSENT STATEMENT**
NAVAL POSTGRADUATE SCHOOL, MONTEREY, CA 93943

Participant:   VOLUNTARY CONSENT TO BE A RESEARCH PARTICIPANT IN:
Testing and evaluation of Iraqi Enrollment via Voice Authentication Project (IEVAP) in support of banking applications in Iraq.

1. I have read, understand and been provided "Information for Participants" that provides the details of the below acknowledgments.

2. I understand that this project involves research.  An explanation of the purposes of the research, a description of procedures to be used, identification of experimental procedures, and the extended duration of my participation have been provided to me.

3. I understand that this project does not involve more than minimal risk.  I have been informed of any reasonably foreseeable risks or discomforts to me.

4. I have been informed of any benefits to me or to others that may reasonably be expected from the research.

5. I have signed a statement describing the extent to which confidentiality of records identifying me will be maintained.

6. I have been informed of any compensation and/or medical treatments available if injury occurs and is so, what they consist of, or where further information may be obtained.

7. I understand that my participation in this project is voluntary; refusal to participate will involve no penalty or loss of benefits to which I am otherwise entitled.  I also understand that I may discontinue participation at any time without penalty or loss of benefits to which I am otherwise entitled.

8. I understand that the individuals to contact should I need answers to pertinent questions about the research are Maj Jeff Withee or Capt Ed Pena, Principal Investigators, and about my rights as a research participant or concerning a research related injury is Prof_____, _____ Dept. Chairperson.  A full and responsive discussion of the elements of this project and my consent has taken place.   **NPS Medical Advisor:** LTC Eric Morgan, MC, USA, Commanding Officer, Presidio of Monterey Medical Clinic, (831) 242-7550, eric.morgan@nw.amedd.army.mil

_____

Signature of Principal Investigator               Date

_____

Signature of Volunteer                            Date

## PARTICIPANT CONSENT FORM

1. **Introduction.** You are invited to participate in a study on the demonstration of the Iraqi Arabic Interactive Voice Response System. With information gathered from you and other participants, we hope to demonstrate the use of an Iraqi Arabic voice-activated menu-driven phone system using existing COTS interactive voice response (IVR) technology in order to expedite a visitor's entry to a controlled facility/secure space. We ask you to read and sign this form indicating that you agree to be in the study. Please ask any questions you may have before signing.

2. **Background Information.** The Naval Postgraduate School's Voice Authentication Technology Research Group is conducting this study.

3. **Procedures.** If you agree to participate in this study, the researcher will explain the tasks in detail. There will be 10 required sessions with each session lasting 1-2 minutes: User will test and evaluate the proof-of concept system by calling into IVR phone system during which you will be expected to accomplish a number of tasks related to appointment scheduling using your Arabic language capabilities.

4. **Risks and Benefits.** This research involves no risks. The benefits to the participants are gaining techniques for the demonstration of this technology for subsequent research and development.

5. **Confidentiality.** The records of this study will be kept confidential. No information will be publicly accessible which could identify you as a participant.

6. **Voluntary Nature of the Study.** If you agree to participate, you are free to withdraw from the study at any time without prejudice. You will be provided a copy of this form for your records.

7. **Points of Contact.** If you have any further questions or comments after the completion of the study, you may contact the research supervisor.

8. **Statement of Consent.** I have read the above information. I have asked all questions and have had my questions answered. I agree to participate in this study.


------------------------------------------ ------------------
Participant's Signature Date


------------------------------------------ ------------------
Researcher's Signature Date

PRIVACY ACT STATMENT

**NAVAL POSTGRADUATE SCHOOL, MONTEREY, CA 93943**

**PRIVACY ACT STATEMENT**

1. Purpose: The purpose of this research is to create a pilot system using existing commercial off the shelf (COTS) technologies in order to help develop the Iraqi banking system. This system will serve as a proof-of-concept (POC) system in the demonstration and pilot evaluation of an Iraqi Arabic voice-activated menu-driven phone system using existing COTS interactive voice response (IVR) technology in order to verify the identity of a user in order to allow them access to a bank's other applications.

2. Use: Data collected from this research will be used for statistical analysis by the Departments of the Navy and Defense, and other U.S. Government agencies, provided this use is compatible with the purpose for which the information was collected. Use of the information may be granted to legitimate non-government agencies or individuals by the Naval Postgraduate School in accordance with the provisions of the Freedom of Information Act.

    1. Disclosure/Confidentiality:
        a. I have been assured that my privacy will be safeguarded. I will be assigned a control or code number which thereafter will be the only identifying entry on any of the research records. The Principal Investigator will maintain the cross-reference between name and control number. It will be decoded only when beneficial to me or if some circumstances, which is not apparent at this time, would make it clear that decoding would enhance the value of the research data. In all cases, the provisions of the Privacy Act Statement will be honored.
        b. I understand that a record of the information contained in this Consent Statement or derived from the experiment described herein will be retained permanently at the Naval Postgraduate School or by higher authority. I voluntarily agree to its disclosure to agencies or individuals indicated in paragraph 3 and I have been informed that failure to agree to such disclosure may negate the purpose for which the experiment was conducted.
        c. I also understand that disclosure of the requested information, including my Social Security Number, is voluntary.

_____

Name, Grade/Rank (if applicable) DOB SSN
[Please print]

_____

Signature of Volunteer      Date

THIS PAGE INTENTIONALLY LEFT BLANK

# APPENDIX C

**EMPLOYMENT AGREEMENT WITH INDEPENDENT CONTRACTOR: general form**

Contract made *May 3, 2007* between Iraqi Voice Enrollment and Authentication Project (IVEAP) of Monterey, CA, here referred to as owner, and _____ *[name]*, of _____ *[city* _____ *[state],,* here referred to as contractor.

## RECITALS

A. Owner is conducting an experiment designed to test the accuracy of a voice authentication program for use in security applications for banking.

B. The contractor agrees to make the number of calls in the manner prescribed below during the time designated in this contract.

In consideration of the mutual promises set forth in this contract, it is agreed by and between owner and contractor:

## SECTION ONE.

## DESCRIPTION OF WORK

The contractor will make a total of _____ calls to be distributed between the first three weeks of the experiment, with a one week break in between the first and third week. During the final week of the experiment the caller will make _____ calls in order to try and defeat the system and try to gain access to someone else's account. The contractor will be designated as a wireless or land-line user (circle one) and understands that the majority of their calls should be of the kind they were assigned. If chosen as part of the advanced imposter trials, the contractor will perform additional trials and be compensated at the prescribed overtime rate.

## SECTION TWO.

## PAYMENT

Owner will pay contractor their current overtime rate depending on the following schedule:
    4 Hours overtime for land line users
    5 Hours overtime for wireless users
    6+ Hours overtime for advanced imposters

SECTION THREE.

Experiment

The contractor agrees that he/she will conduct the experiment as explained to them by the primary investigators. Specifically, they will not use speakerphones or hands-free phone systems as this will affect the quality of their voice print and skew the results of the test. Other than the aforementioned restriction, the contractors are encouraged to call in to the system at different times of the day and in different environments. If they are not successful at accessing their account, they should record the time of the call and the mitigating factors (background noise, bad signal etc.) and report them to one of the principal investigators (edpena@nps.edu or jwwithee@nps.edu).

DURATION

Either party may cancel this contract on one week's written notice; otherwise, the contract shall remain in force for a term of two months from the date the contract is signed or until the experiment is completed (whichever occurs first).

In witness whereof, the parties have executed this agreement at *the Defense Language Institute, Monterey CA* the day and year first above written.

_____
Volunteer

_____
Investigator

# APPENDIX D

Schedule:
07-13 May:
- One Group will strictly use cell phones
- One Group will strictly use standard Telephones
- Enrollment and Verifications (24 Hours a day Seven days a week)
- Fifteen Verifications

14-20 May:
- Break

21-27 May:
- Enrollment and Verifications (24 Hours a day seven days a week)
- Fifteen Verifications

28 May – 3 June:
- Imposter Trials
- Thirty Verifications

Telephone Number:
(831) 656 1912

Enrollment:
- Say or Key in your account number.
  - It only recognizes single digits: e.g..: "one two three four five six seven eight nine zero"
  - Notice that if an account number has already been enrolled, it will go through the "verification" dialog below, not through this one.
- I heard "one two three four…" did I get that right?
  - Say yes if the account number was right. Say no otherwise.
- Now, it looks like we have not yet enrolled you in…. otherwise to go ahead with the enrollment process right now say "enroll me now"…
  - Say "enroll me now"
- Now, to create your voice print I will ask you to count out loud from 1-9…
  - Say "one two three four five six seven eight nine"
- And once more please:
  - Say "one two three four five six seven eight nine"
- And one last time:
  - Say "one two three four five six seven eight nine"
- Ok, your voice is enrolled, so everything is setup for your next call.

<u>Verification:</u>
- Say or Key in your account number.
  - It only recognizes single digits: e.g..: "one two three four five six seven eight nine zero"
  - Note: The account number must have been enrolled before, otherwise it will go through the dialog sequence above.
- I heard "one two three four…" did I get that right?
  - Say yes if the account number was right, otherwise say no.
- Now to verify your voice please count out loud from 1 up to 9
  - Say "one two three four five six seven eight nine"
- The system might ask again to repeat 1-9.
  - Say "one two three four five six seven eight nine"
- The system will accept the user by saying: "you've been verified…"
- Or the system will reject the user by saying: "I am sorry I am having trouble verifying your account information…"

<u>Imposter Trials:</u>
-After completion of the Verifications, you will be e-mailed a list of thirty different account to try and access.
-The procedures are the same as the Verification (with the exception that you should be rejected not accepted)
-If you are accepted by an account try the account again to see if you were able to gain access a second time.
-E-mail Major Jeff Withee at jwwithee@nps.edu with the account numbers you were able to access and whether or not you were able to access that again a second time.

<u>Usability:</u>
After completing the imposter trials, please let us know a yes or no on whether or not you felt the system was easy to use.

<u>A couple of recommendations:</u>
- It's better to use "sahia" (Iraqi Arabic for correct) instead of na'am (Iraqi Arabic for "yes").
- At the time of enrollment, you can use any sequence of digits for PIN number.
- It's better to get very familiar with the system dialog flow as unfamiliarity with a dialog flow is the number 1 cause of errors in demonstrations.
-Please do not use speakers or hands-free devices.
-Try to be a place where there is very little background noise.

Please, do not hesitate to call one of us if you need to know anything:
Atheer: (831) 242-6908 (Arabic Speaker)
Eddie: (831) 917-0073
Jeff:   (760) 207-9639

Thanks!

# LIST OF REFERENCES

[1]     Special Inspector General for Iraq Reconstruction.  "Quarterly Report to Congress", www.sigr.mil, April 30, 2007. [Online]. Available: http://www.sigir.mil/reports/quarterlyreports/Apr07/pdf/Report_-_April_2007_Complete.pdf. [Accessed: July 20, 2007].

[2]     D. Petraeus.  "No military solution to Iraq, U.S. general says", March 9, 2007. [Online]. Available: http://www.cnn.com/2007/WORLD/meast/03/08/iraq.petraeus/index.html [Accessed: July 25, 2007].

[3]      D.S. Alberts, J.J. Garstka, and F.P. Stein, *Network Centric Warfare: Developing and Leveraging Information Superiority*, 2nd Edition.  Washington DC: CCRP Publication Series, 2000, p. 6.

[4]     Liu, S. and Silverman, M., "A Practical Guide to Biometric Security Technology," IEEE Computer Society, IT Pro - Security, January-February 2001.

[5]     Tech Target Inc. "Definition of Nonrepudiation," Tech Target Inc. [online]. Available: http://searchsecurity.techtarget.com. [Accessed: August 29, 2007].

[6]     B. Gold, and N. Morgan, *Speech and Audio Signal Processing:  Processing and Perception of Speech and Music*. John Wiley & Sons, Inc New York 2000.

[7]     Proof of Concept: Iraqi Enrollment Via Voice Authentication Project, NPS Master's Thesis, Samuel Lee, September 2005.

[8]     R.L. Klevans, and R.D. Rodman, *Voice Recognition*, Artech House, 1997.

[9]     Lumenex Inc. "Voice Biometrics Solutions," Lumenex Inc. [Online]. Available: http://www.lumenex.com/Products/VoiceBiometrics.htm#Verification. [Accessed: August 29, 2007].

[10]    U.K. Biometrics Working Group. *Use of Biometrics for Identification and Authentication Advice on Product Selection* Issue 2.0 OeE:   Biometrics Working Group, March 2002.

[11]    S. K. Dahel and Q. Xiao. "Accuracy Performance Analysis of Multimodal Biometrics," Proceedings of the 2003 IEEE Workshop on Information Assurance United States Military Academy, West Point, NY, pp.170-173, June 2003.

[12]    Phase 1C Iraqi Arabic Language Voice Verification Accuracy Estimation Test Plan, NPS, Dr. Pat Sankar, December 2005.

[13]  E-mail correspondence between Dr. Pat Sankar, NPS, and the authors, 23 May 2007.

[14]  Speech Technology Magazine, "Speech Technology Magazine Unveisls the 2006 Speech Solutions Awards", www.speechtechmag.com, August 2006 [Online]. Available: http://www.speechtechmag.com/Articles/PrintArticle.aspx?ArticleID=35530 [Accessed: July 20, 2007].

[15]  Nuance Communications Incorporated, *Homepage*, www.nuance.com, July 2007 [Online]. Available: http://www.nuance.com/company/ [Accessed: July 20, 2007].

[16]  Nuance Communications Incorporated, "Press Kit", www.nuance.com, July 2007 [Online]. Available: http://www.nuance.com/news/presskits/ [Accessed: July 13, 2007]

[17]  Nuance Communications Incorporated, "Nuance Recognizer Datasheet", www.nuance.com, July 2007 [Online]. Available: http://www.nuance.com/recognizer/datasheets/ [Accessed: July 20, 2007].

[18]  Nuance Communications Incorporated, "Nuance Verifier Datasheet", www.nuance.com, July 2007 [Online]. Available: http://www.nuance.com/verifier/datasheets/ [Accessed: July 20, 2007].

[19]  Nuance Communications Incorporated, "Nuance Voice Platform Datasheet", www.nuance.com, July 2007 [Online]. Available: http://www.nuance.com/voiceplatform/datasheets/ [Accessed: July 20, 2007].

[20]  National Security Council, *National Strategy for Victory in Iraq*. Washington, D.C.: National Security Council, 2005.

[21]  Naval Postgraduate School, *Brinkley Brief 07120*. Monterey, CA: NPS, 2007.

[22]  Islamic Republic of Afghanistan, Office of the President, "Infrastructure and Natural Resources," Office of the President. [Online]. Available: http://www.president.gov.af/english/np/naturalresources.mspx#Telecom. [Accessed: August 30, 2007].

[23]  Central Intelligence Agency, World Fact Book: Afghanistan, [Online]. Available: https://www.cia.gov/cia/publications/factbook/geos/af.html#Comm [Accessed: August 30, 2007].

[24]  Nuance, "Managing and Administering the Nuance Voice Platform Student Guide ver 2.1," June 10, 2004.  Menlo Park, CA.

[25]     Erlang Home Page, "Erlang Calculator." [Online]. Available: http://www.erlang.com/calculator/erlb/. [Accessed Aug 30, 2007].

[26]     M. Sauer, "Report: Corruption Investigations in Iraq Stopped in Their Tracks" April 30, 2007 [Online]. Available: http://blogs.abcnews.com/theblotter/2007/04/report_corrupti.html  [Accessed August 30, 2007].

[27]     P.M. Senge, P.M. *The Fifth Discipline: The Art and Practice of the Learning Organization*, New York: Double Day, 1990.

[28]     D.A. Nadler and M.L. Tushman, *The Congruence Model: A Roadmap for Understanding Organizational Performance,* Mercer Delta Publications, 2003. [Online]. Available: http://www.oliverwyman.com/ow/. [Accessed: August 30, 2007].

[29]     M.A. Beer, "Leading Change" in Harvard Business School, Harvard: Publishing Division, 1988.

[30]     J.P. Kotter, Leading Chang: Why Transformation Efforts Fail. HBR March-April, 1995.

[31]     W. Bridges, Managing Transition, Reading, Mass: Addison-Weley, 1991.

[32]     E.H. Schein, Organizational Psychology, 3rd ed. Englewood Cliffs, N.J: Prentice Hall, 1980.

[33]     T.D. Jick, "The Vision Thing (A)," Harvard Business Review. Harvard: Publishing Division, 1989.

THIS PAGE INTENTIONALLY LEFT BLANK

# INITIAL DISTRIBUTION LIST

1.      Defense Technical Information Center
        Ft. Belvoir, Virginia

2.      Dudley Knox Library
        Naval Postgraduate School
        Monterey, California

3.      Marine Corps Representative
        Naval Postgraduate School
        Monterey, California

4.      Director, Training and Education, MCCDC, Code C46
        Quantico, Virginia

5.      Director, Marine Corps Research Center, MCCDC, Code C40RC
        Quantico, Virginia

6.      Marine Corps Tactical Systems Support Activity (Attn: Operations Officer)
        Camp Pendleton, California

7.      Gerald Christman
        Office of the Secretary of Defense
        Pentagon, Washington D.C.

8.      Stephen Nunn
        SPAWAR Systems Center
        Point Loma, California

9.      Dan Boger
        Naval Postgraduate School
        Monterey, California

10.     James F. Ehlert
        Naval Postgraduate School
        Monterey, California

11.     Pat Sankar
        Naval Postgraduate School
        Monterey, California

12. Ramon Prieto
    Nuance Communication, Inc.
    Menlo Park, California